Pushing the boundaries for automated data reconciliation in official statistics

JACCO DAALMANS

Pushing the boundaries for automated data reconciliation in official statistics

Jacobus Adriaan Daalmans

ISBN: 978-94-6361-227-2

Lay-out and printing by Optima Grafische Communicatie

Pushing the boundaries for automated data reconciliation in official statistics

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University
op gezag van de rector magnificus, prof.dr. E.H.L. Aarts,
in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie
in de Aula van de Universiteit op vrijdag 22 maart 2019
om 13.30 uur

door

Jacobus Adriaan Daalmans,

geboren te Groningen.

Promotores:

Prof. dr. A.G. de Waal

Prof. dr. J.K. Vermunt

Promotiecommissie

Prof. dr. J.A. van den Brakel

Prof. dr. P.G.M. van der Heijden

Dr. K. Van Deun

Dr. J.W. van Tongeren

DANKWOORD

Een proefschrift schrijven heb ik lange tijd afgehouden.

Een collega vertelde ooit dat iedere promovendus in een crisis belandt. Daar had ik geen zin in. Jaren later kwam dan toch een voorstel dat ik niet wilde weigeren. Ik mocht promoveren op een onderwerp dat me al tijden bezighield en bovenal had ik de juiste mensen om me heen. Nu mijn proefschrift bijna klaar is, wil ik diegenen bedanken.

Allereerst Ton de Waal: ooit nam je me, na een zenuwachtige sollicitatie, aan als stagiair op het CBS. Uiteindelijk werd je de promotor die ik mezelf toewens.

Van jou heb ik meegekregen dat wetenschap leuk en creatief is. De vrijheid die je me gaf heeft een proefschrift opgeleverd waar ik trots op ben. De onderwerpen van dit proefschrift zijn bijna vanzelf opgekomen en vormen toch een mooi samenhangend geheel. Je nonchalant uitgesproken “ik weet het niet” stimuleert tot zelf nadenken. Het meest typerend van jou is je humor die altijd de nodige lucht(igheid) geeft. “Je rechter veter zit vast”, om maar eens wat te noemen. Ton, ik hoop nog lang met je samen te werken.

Professor Jeroen Vermunt, met jou had ik alleen aan het eind contact, je vakgebied is heel anders dan dat van mij, maar ik heb zeker geprofiteerd van je uitgebreide ervaring en je concrete, nuttige adviezen. Ik ben vereerd dat jij mijn promotor wilde zijn.

Graag wil ik ook de leden van de leescommissie bedanken voor het lezen en beoordelen van het manuscript. Prof. dr. J.A. van den Brakel, Prof. dr. P.G.M. van der Heijden, Dr. K. Van Deun en Dr. J.W. van Tongeren, dank voor jullie tijd!

Ook wil ik veel directe CBS-collega’s bedanken.

Reinier Bikker en Nino Mushkudiani: het resultaat van onze samenwerking is in veel hoofdstukken te zien. Dankzij jullie is er meer aandacht voor macro integratie, niet alleen op het CBS, maar ook daarbuiten. Het schrijven van dit proefschrift werd daardoor een stuk makkelijker.

Directe collega’s: Jeroen Pannekoek, Paul Knottnerus, Sander Scholtus, Harm Jan Boonstra, Mark van der Loo, Edwin de Jonge, Arnout van Delden en Laura Boeschoten: ik heb niet alleen inhoudelijk veel van jullie geleerd, maar waardeer ook jullie vriendelijkheid, waardoor ik geen enkele schroom voel om van jullie kennis gebruik te maken.

Overige methodologen en procesontwikkelaars: ik werk altijd graag met jullie. Niet alleen met de Haagse collega’s, maar zeker ook met de Heerlenaren. Regelmatig krijg ik de vraag van een Heerlense collega wanneer ik nu eindelijk eens naar het zuiden verhuis. Hoewel ik die vraag inmiddels wel ken, zorgt die ervoor dat ik me welkom voel.

Ik wil graag mijn leidinggevende Bart Bakker bedanken voor de gelegenheid die hij bood om dit proefschrift te schrijven. Oud-leidinggevende Piet van Dosselaar: bedankt voor de introductie met macro integratie en het geduld dat je me gaf om me in het onderwerp te verdiepen.

Verder wil ik verschillende collega’s van andere afdelingen van het CBS bedanken.

Brugt Kazemier: voor de stimulerende kennismaking met Nationale Rekeningen.

Jan van Dalen, Coen Leentvaar, Harm Melief, Vincent Ohm, Marcel Pommée, Willem Sluis en Ronald van der Stegen: voor de ontwikkeling van de 'inpasmachines'.

Eric Schulte Nordholt en Frank Linder: voor de levendige samenwerking bij de Volkstelling.

Danny van Elswijk: voor een toepassing bij bedrijfseconomische statistieken, die wat lastiger bleek dan verwacht. De inzichten die dat opleverde zijn de directe aanleiding voor hoofdstuk 4.

Daarnaast zijn er nog veel andere CBS'ers met wie ik met plezier heb gewerkt. In het bijzonder dankbaar ben ik collega's van inhoudelijke afdelingen die bereid zijn nieuwe methoden een kans te geven in 'hun' productieproces.

In de afgelopen periode heb ik verschillende internationale congressen bezocht. Twee mensen daar wil ik in het bijzonder bedanken.

Prof. di Fonzo, dear Tommy, you are definitely the most passionate and gentleman-like researcher I know. Before I met you in person, i looked up to you as the author of many important papers. Later, we started working together. This has been a very valuable experience to me. It has been a great honour to present our joint works in Pisa and Paris. I am pleased that you are a co-author of Chapter 3. Your emails often make me smile and always make me feel good. Tommy, you are an example to me in many ways.

Baoline Chen: It has been a great pleasure to know you. Thank you very much for our discussions in different meetings. Your editing of my manuscript for the special edition of Statistica Neerlandica and Chapter 4 of this thesis has been highly appreciated.

Tot slot, wil ik vrienden en familie bedanken.

Allereerst Weekendhike, waar ik veel leuke mensen heb ontmoet. Het voert te ver iedereen persoonlijk te bedanken.

Beerhikers proost! Carl, Frank, Guus, Henk, Peter en Richard, we hebben menig bierfestival bezocht en daar van alles gedronken tot aan wc-reinigers aan toe.

Maaike bedankt voor de vele gezelligheid en positiviteit. Na een spelletje Wordfeud met jou sta ik altijd weer met beide benen op de grond.

Schoonfamilie, Annie, Bert en Sonja Staals bedankt voor het medeleven en de afleiding die jullie bieden.

Mijn ouders, Peter en Marieke Daalmans, bedankt! Jullie huis voelt altijd als thuis. Jullie leerden me hoe belangrijk geduld is en stimuleerden me altijd om door te leren.

Peter, ik kan me herinneren dat je lange tijd geleden vertelde dat je verwachtte dat ik eens zou promoveren, nog voordat ik zelf concrete plannen had. Ik ben blij dat je zo als paranimf achter me staat.

Oma, wat voel ik me bijzonder dat ik een oma heb.

Margreet, Michel, Thijs en Robin, het is altijd fijn om jullie te zien en helemaal geweldig om tikkertje, verstoppertje of een ander onbegrijpelijk spel met jullie te spelen. Daar word ik altijd blij van.

Lieve Julie, maatje, rustpunt en steun. Door jou sta ik er niet alleen voor. Jouw gedrevenheid werkt aanstekelijk. Je hulp bij het organiseren van de verdediging is onmisbaar. Bovenal ben je mijn lieve vriendin.

CONTENTS

1. Introduction 11

2. Benchmarking large accounting frameworks: a generalised multivariate model 21

3. GRP temporal benchmarking: drawbacks and alternative solutions 45

4. On the sequential benchmarking of sub-annual series to annual totals 65

5. Divide-and-Conquer solutions for estimating large consistent table sets 87

6. Constraint simplification for data editing of numerical variables 107

7. Discussion 123

8. References 131

9. Summary 141

1. Introduction

National Statistic Institutes (NSIs), such as Statistics Netherlands, have to publish reliable and coherent statistical information. To meet this requirement, estimates of the same phenomenon based on different data sources should ideally be the same. An example of a possible inconsistency that should be prevented from occurring is that the amount of bottles of wine sold to consumers is not the same, depending on whether the result is observed from sellers or buyers of these bottles. Inconsistent results are undesirable as they cause uncertainty. Numerical consistency of statistical results does not naturally happen. In the previous example, it is well known that people tend to underreport alcohol use, meaning that wine consumers tend to report a lower amount than sellers. Besides measurement error, various other causes for discrepancy exist, such as sampling error, nonresponse error, coverage error and processing error (Eurostat, 2009). As mentioned by Di Fonzo and Marini (2005), data are often incomplete at some level of disaggregation. Inconsistency cannot only be observed for one statistical output, but more generally, it can also refer to relations between variables, e.g. profits which need to be the same as the difference between turnover and cost. To detect inconsistencies in statistical data a framework is needed that includes a set of definitions and relations between variables. Statistical tables are said to be numerical consistent if the data satisfy a set of predefined relations.

A typical example of such a framework is the national accounts (NA). The NA include key economic indicators, of which Gross Domestic Product (GDP) is the most well-known. National accounts have been published in the western world since the 1940s. The Dutch economist Jan Tinbergen was an important pioneer in the development of econometric models for national accounts compilation. He received the first Nobel Prize for economics in 1969. Another important contribution came from Sir Richard Stone. The principle of "double accounting" can be attributed to him, stating that every item on one side of a balance must be met by an item on the other side. Stone won a Nobel prize in Economic sciences in 1984. Many accounting rules are defined for NA tables. An example, directly stemming from Keynesian theory, is that for any commodity in the economy total supply must match total use. Total supply includes production and imports. Total use comprises (inter-mediate) consumption, investments, stocks adjustments and exports. For instance, the amount of money farmers receive for producing cucumbers should match the amount spent on cucumbers by consumers, companies and the government. Data for NA tables are fed by different kinds of independent sources that greatly vary in accuracy. These sources can be surveys that are conducted by statistical institutes, but also register data obtained from public administration or even expert guesses. Because of the different kinds of errors that have been previously mentioned, the data that have been compiled from these sources usually do not satisfy the consistency rules.

Macro integration is the term which is typically used when referring to the process to be followed to arrive at consistent results. It involves adjusting the preliminary values at an aggregate level. A main distinction can be made between corrections for bias and random errors. Bias refers to deviations that are not due to chance alone. It often relates to an error with a known cause, whose expected value structurally differs from zero. Underreporting of wine consumption is an example of this. Random errors, on the other hand, appear more or less by accident, often due to sampling error. The expected value is zero, meaning that on average the error will be close to zero if measurement could be repeated. Macro integration first cleans the data for bias and then solves the remaining, often smaller, random discrepancies. This thesis solely focuses on the last step, the correction of random errors.

NSIs have often applied informal methods for macro integration, in particular for National Accounts reconciliation. Such methods rely on agreement among subject matter experts on the adjustments to be made to the raw data or the obtained tables. Despite that informal methods have been working well, they also have drawbacks. One of these is that the process is not transparent and thus irreproducible. This is why Kooiman *et al.* (2003) refer to these methods as "voodoo". Another drawback is its time-consuming nature. National Account tables are often very detailed, consisting of thousands of cells. Because of the many relations between different variables, a change of one value might imply a need to adjust several other cells. The reconciliation process is a challenging puzzle that requires extensive knowledge about the economy and the relations between the different variables.

A wide range of formal macro integration methods is available in the literature that can be used as an alternative for the previously mentioned informal methods. A distinction can be made between methods for one-period data and methods for time series data. One of the most prominent methods in the former category is the least-squares adjustment method by Stone *et al.* (1942). A method, especially useful for time series data, is Denton's method (Denton, 1971).

This thesis' aim is to push the boundaries of automated macro integration methods for compiling official statistics. Because of their importance for the further chapters, Stone's method and Denton's method are explained in Sections 1.1 and 1.2. Section 1.3 briefly explains the close relation between macro integration and data editing, i.e. the problem of removing inconsistencies from data at the micro level. The reason for including this section is that one of the chapters of this thesis, Chapter 6, deals with a data editing problem, which also has relevance for macro integration. Thereafter, Section 1.4 summarizes problems and opportunities of macro integration methods. Finally, Section 1.5 provides an outline of the remainder of this thesis.

1.1 STONE'S METHOD

The idea of most formal macro integration methods is to find a solution for the integration problem with the least possible disruption of the source data. Stone's method uses a quadratic loss function, which is very well known in statistics. In principle, all data might be adjusted, but differences in reliability can be taken into account. Items that are known to be reliable are usually designated to be adjusted less than 'unreliable' items.

Stone's method postulates that observed items can be represented as a vector $\boldsymbol{x}$ that can be written as a sum of latent 'true' values $\boldsymbol{x}_0$ and an error $\boldsymbol{\epsilon}$ with zero mean and a known covariance matrix $\mathbf{V}$. The reconciled values $\boldsymbol{x}^*$ need to obey a set of linear constraints given by

$$\mathbf{A}\boldsymbol{x}^* = \boldsymbol{b}. \tag{1.1}$$

The reconciled values in $\boldsymbol{x}^*$ are obtained by minimising

$$(\boldsymbol{x}^* - \boldsymbol{x})' \mathbf{V}^{-1} (\boldsymbol{x}^* - \boldsymbol{x}) \tag{1.2}$$

subject to the constraints in (1.1), a problem that belongs to the class of convex quadratic optimization problems (QP).

The optimal solution for $\boldsymbol{x}^*$ and the corresponding covariance matrix $\mathbf{V}^*$ are given by

$$\begin{aligned} \boldsymbol{x}^* &= \boldsymbol{x} - \mathbf{VA}'(\mathbf{AVA}')^{-1}(\mathbf{A}\boldsymbol{x} - \boldsymbol{b}) \quad \text{and} \\ \mathbf{V}^* &= \mathbf{V} - \mathbf{VA}'(\mathbf{AVA}')^{-1}\mathbf{AV}. \end{aligned}$$

One can mathematically prove that the variances in $\mathbf{V}^*$ are no larger than the variances in $\mathbf{V}$, formally showing that data reconciliation improves accuracy. The solution of Stone has certain attractive mathematical properties; it is for instance an unbiased estimator, and of all unbiased estimators it is the most accurate one.

At the time Stone's method was devised, it could not be applied to large macro integration problems. A main complication is the computation of the inverse $(\mathbf{AVA}')^{-1}$. This computation is resource-intensive. Byron (1978) proposed the use of a numerical iterative method to solve the quadratic optimisation problem, a so-called conjugate gradient method, which avoids matrix inversion. The improved performance of Byron's method comes at the cost that the covariance matrix $\mathbf{V}^*$ cannot be computed. But for many applications this covariance matrix is not needed. Today, quadratic optimisation problems are well-studied and efficient numerical methods have been implemented in commercial and open source software. Modern software can easily deal with very large problems consisting of 100,000 variables and even more. This thesis relies on these methods and hence directly builds on the ideas of Byron (1978). Since the possibilities of automated data integration have been dramatically improved, problems that were not eligible for automated methods in the past can now be easily processed.

A second complication of Stone's method is the limited modelling options. The model defined in (1.1) and (1.2) takes account of linear 'equality' constraints only. However for many real-life applications, constraints that cannot be directly formulated in this form have to be inevitably considered. For example, it often occurs that a ratio of two variables should have a certain known value. An example is that the value added tax paid by an economic industry has to be a fixed percentage of the output of that industry. Also frequently occurring is the constraint that economic variables cannot have a negative value. Moreover, certain "soft" relations have to be taken into account, i.e. relations that only need to hold by approximation. One could for instance expect a relation between the production of milk and cheese. For the production of one kilogram of cheese a certain amount of milk is needed, an amount which can be expected to be quite stable over time. Therefore, it is unlikely, but not yet impossible, that an increase of cheese production goes together with a substantially lower (intermediate) use of milk. Several works in the literature have extended Stone's method to allow for a larger class of constraints. Magnus *et al.* (2000), for example, developed a Bayesian method that is suitable of including all of the previously mentioned examples.

1.2 DENTON'S METHODS

Many statistical institutes present statistical output on the same variable at different time intervals, e.g. quarterly and annually. Sometimes it is required that a certain 'temporal' aggregation relation is fulfilled, for instance that four quarterly values sum up to one annual value. If quarterly and annual data are independently produced consistency is not automatically achieved. The process of achieving consistency in statistics based on time series is called benchmarking. The main principle of most benchmarking methods is that low frequency data are fixed, since these are usually based on the most comprehensive data sources. Consequently, the high frequency data need to be adjusted. In doing this, one often tries to maximally preserve the one period-ahead movements. The reason is that high frequency series are especially geared towards measuring short-term change. A well-known class of benchmarking methods are the ones proposed by Denton (Denton, 1971). Mathematically, Denton's methods belongs to the family of least-squares adjustment methods, just like Stone's method. Denton is very popularly applied, mainly due to its simplicity. The Denton methods were originally developed for univariate series, but many extensions have been described in the literature. Some of these consider the multivariate case, in which several time series need to be simultaneously processed and relations between series must hold at each time period.

1.3 RELATION BETWEEN MACRO INTEGRATION AND DATA EDITING

The data editing problem is closely related to macro integration, but yet different. Because one the chapters in the remainder of this thesis (Chapter 6) is of interest to both problems, we briefly describe their relation below.

Where macro integration achieves consistency between data from different sources, data editing deals with inconsistencies within the records of a single data source. Data editing is not a macro integration method, because corrections are made at the level of the individual respondent. The aim of data editing is to find inconsistencies in the data provided by respondents. A classic example is a male who reports to be pregnant. Data editing can be divided into error localisation, the problem of identifying the erroneous fields of a record, and imputation, the process of filling in plausible values for erroneous or missing values. Error localisation is often done according to the Fellegi-Holt paradigm (Fellegi and Holt, 1976), stating that a minimum set of values need to be identified that can be corrected such that a consistent record is achieved. The underlying assumption is that most of the answers given by a respondent are correct. Compared to macro integration, the emphasis of data editing is more on finding errors than on cleaning the data for small disturbances. Hence, both problems have a different goal. Most macro integration methods try to minimize total adjustment, where the number of adjusted values does not matter. For data editing this is the other way around: it usually attempts to minimize total number of corrections, whereas the size of corrections is irrelevant. Mathematically, the error localisation problem translates into a mixed integer programming (MIP) problem; which is closely related, but somewhat more difficult than the standard quadratic programming problems that are obtained for most macro integration problems.

1.4 PROBLEMS AND OPPORTUNITIES

When implementing formal macro integration methods, several challenges arise that have not been completely solved so far. This thesis provides solutions for some complications that have been actually faced by Statistics Netherlands. Moreover, the rapid development of macro integration methods also provides opportunities; macro integration methods may avoid problems that are faced with other techniques. This thesis first discusses three complications that relate to benchmarking. Then, the merits of Stone's macro integration method are investigated for a problem in the field of social statistics. Finally, attention is paid to constraints handling, a problem that arises after the implementation of any data integration method. The above-mentioned topics are briefly explained below.

A first complication of most benchmarking methods is that these methods have rather restrictive assumptions. Most methods only allow for linear constraints. In practice, there is a need to deal with sophisticated relations between economic variables. Hence, there is a need for support of a broad class of constraints. In the literature, several extensions for Stone's method are already available that enable sophisticated modelling constructions, see Subsection 1.2. Similar extensions for benchmarking methods used to be missing.

A second complication is the choice of an appropriate benchmarking method. Denton's methods are very popularly applied, mainly because of their simplicity. Some works in the literature argue however that the so-called Growth Rate Preservation (GRP) method should be preferred, because of better theoretical foundations. Although a few comparison studies are available, an in-depth comparison of these two approaches has not been conducted so far.

A third complication of benchmarking methods is that when benchmarking time series, theoretically, it is best to use all available data of the past, but in practice, such an approach is not always feasible. The reason is that Denton leads to new results for a whole time series. This can be problematic, because in practical applications, as those of NSIs, results of the past may not be allowed to change, for instance, because these have already been published. Hence, benchmarking is often applied to relatively short time-series. A drawback of this is that abrupt breaks in benchmarking corrections might be observed between sequentially estimated series. These breaks do not comply with the aim of benchmarking.

Besides challenges macro integration methods also provide opportunities. There is a growing tendency to benefit from macro integration methods outside the field of National Accounts. A potential new application area is the Dutch Population census. For this application, many detailed tables have to be estimated from different data sources. Numerical consistency is a key requirement. The two latest censuses were produced by using a weighting method. Several estimation problems were experienced by the application to the detailed census tables. The application of macro integration techniques might solve these problems.

A further complication arises after the implementation of any data integration method. This complication amounts to the problem of setting up and maintaining a set of constraints. In a practical setting, the number of constraints may be very large. Imposing many restrictions between variables might pose several problems: low software performance, errors in rule formulation that remain hidden in the bulk of rules and poor insight in interdependency of the many rules. This thesis considers to use constraint simplification techniques to solve these problems.

1.5 THESIS OUTLINE

The main body of this thesis consists of five published journal articles that extend the current knowledge on the theory and practice of macro integration methods. These chapters connect to the problems and opportunities as identified in Section 1.4. Since the five chapters can be independently read, some overlap in the text occurs and some inconsistency in notation may be observed across chapters. A short overview of the chapters is given below.

Chapter 2 presents a new multivariate Denton method that is currently applied in the production process of Dutch National Accounts. The new method extends an already existing multivariate Denton method (Di Fonzo and Marini, 2003) with new modelling features that have been originally proposed for Stone's methodology (Magnus *et al.*, 2000). Chapter 3 enriches the current knowledge about the differences between Denton and the Growth Rate Preservation (GRP) methods. Chapter 4 proposes solutions for the 'sequential estimation' problem by avoiding large steps between sequentially estimated series. Chapter 5 examines the use of a Stone-based method for the Dutch Population census. Finally, Chapter 6 describes the problem of setting up and maintaining a large set of constraints.

2. Benchmarking large accounting frameworks: a generalised multivariate model[1]

Summary. We present a multivariate benchmarking model for achieving consistency between large quarterly and annual accounting frameworks. The method is based on a quadratic optimization problem, for which many efficient numeric solvers exist. The method combines several features, such as linear constraints, ratio constraints, weights, and inequalities, in one model. Therefore a wide range of modelling possibilities is supported. This method is especially interesting for national statistical offices, to simplify their processes to achieve consistency between publications.

1 This chapter has been published as Bikker, R.P., J.A. Daalmans and N. Mushkudiani, (2013) Benchmarking large accounting frameworks: a generalized multivariate model. *Economic Systems Research,* 25, 390-408.

2.1 INTRODUCTION

Macro integration is the process for achieving consistency between economic data. By combining data sources, more information is used, yielding more accurate statistics (Boonstra *et al.*, 2010). A problem that often arises while compiling National Accounts is the inconsistency in the source data. Discrepancies are caused by various kinds of errors, like sampling error, non-response error, coverage error, measurement error and processing error (Federal Committee on Statistical Methodology, 2001).

The first step of macro integration consists of correcting errors, in which large obvious discrepancies are detected and corrected. The second step is a reconciliation process; in this step data are corrected so that certain accounting constraints are being fulfilled. The literature on data reconciliation goes back to Stone *et al.* (1942), who presented a constrained, generalised least squares method. Several other reconciliation methods are described in Wroe *et al.* (1999, Annex A).

This chapter focuses on a special case of the reconciliation problem, called benchmarking. Benchmarking achieves consistency between low- and high- frequency time series. Without loss of generality, it is assumed here that the high frequency data are quarterly figures and that these have to be aligned with annual benchmarks. Typically, the annual data sources provide the most reliable information about overall levels and the quarterly data sources provide information about short-term changes. In general, the annual data are fixed for this reason.

Benchmarking methods can be broadly classified into purely numerical methods and model-based methods. Bloem *et al.* (2001, Chapter VI) and Dagum and Cholette (2006) give a comprehensive overview of these methods.

The model-based class of methods encompasses regression models (see Cholette-Dagum, 1994), ARIMA model-based methods (e.g. Hillmer and Trabelsi, 1987) and state space models (e.g. Durbin and Quenneville, 1997). Closely related to the regression method is the method of Chow and Lin (1971). Here, the authors derive quarterly data from annual data by using indicator time series, although their method is not a benchmarking method in the strict sense. The Chow and Lin method may suffer from step problems, i.e. large gaps between the fourth quarter of one year and the first quarter of the next year. A modification by Fernández (1981) corrects for this step problem. Rossi (1982) and Di Fonzo (1990) extended the regression method for the multivariate case.

A classical reference to a numerical method is Denton (1971). The Denton method is a quadratic programming method that was initially proposed for univariate data. The aim of this method is to make the quarterly data coherent with annual totals, while preserving all quarter-to-quarter changes as much as possible, the so-called movement preservation principle. Because of this property the Denton method avoids the step problem.

Di Fonzo and Marini (2003) have extended the Denton method for multivariate data. In addition to temporal alignment, multivariate data often also have to satisfy a set of constraints between different variables within the same time-period. Subsequently, Bikker and Buijtenhek (2006) have added reliability weights to the multivariate Denton method.

Although the Denton method is different from the model based methods, under certain conditions both lead to the same results (Fernandez, 1981). An advantage of the model-based methods over the quadratic programming approach is that measures of accuracy e.g. covariance matrix of the benchmarked data can be derived. On the other hand, as mentioned by Bloem *et al.* (2001), the Denton method is very well suited for large scale applications as it is based on the Euclidian norm and linear constraints.

The benchmarking method described in this chapter is based on the multivariate method of Bikker and Buijtenhek (2006). In order to incorporate economic relations in the model, specifically for the National Accounts, we added extra methodological features. These are: soft constraints, ratio constraints and inequality constraints. We adopted the same approach as Magnus *et al.* (2000), who included these features, with the exception of inequality constraints, into a reconciliation method, although their method is not directly intended for benchmarking purposes.

In 2010 Statistics Netherlands implemented this multivariate benchmarking method in its production process of Dutch supply and use tables. For this application very large data sets have to be handled, i.e. over 10 000 time series. For this reason we chose a multivariate Denton method.

Based on a state-of-the-art, commercial quadratic programming (QP) solver XPRESS, we developed a software application tool. The US Bureau of Economic Analysis uses similar software for the implementation of a reconciliation method (Chen, 2006), but their tool is not built for benchmarking. To our best knowledge, benchmarking methods for large data sets have not been applied at other national statistical institutes.

Twenty years ago it was less attractive to implement benchmarking software, since reconciling large disaggregated accounting systems imposes large demands on software capability and especially computer memory. The vast increase in computer power and the development of highly efficient optimization algorithms has dramatically increased the applicability of automatic benchmarking procedures. Problems that were too large to solve on a mainframe computer in the 1980s are easily solved today on a desktop computer.

The introduction of the new benchmarking method led to a large gain in efficiency, when compared to the prior informal methods used by Statistics Netherlands, see e.g. Bloem *et al.* (2001, Chapter VI). Firstly, a formal method yields the same results for the same input of the model. Secondly, the model is built on a firm statistical basis: the adjustments to the data are inversely proportional to the square of the reliability of the data. Thirdly, the least square framework minimises the data adjustment according to the

Euclidean norm which represents our intuitive understanding of "smallest adjustments". Fourthly, the model is flexible: variables and constraints can be easily added or removed.

Compared to other formal methods in the literature, our Denton method is the only method that satisfies all needs of Statistics Netherlands:

- It combines the proportional Denton method with the additive method in one model;
- It is suitable for an application to very large multivariate data sets (500 000 records and more);
- It offers a wide range of possibilities of incorporating relationships into the model, by using hard and soft constraints, equality and inequality constraints and reliability weights;
- It allows for missing data. In particular, annual totals do not have to be available for each year of each time series;
- It allows for time series of different length within a single multivariate model;
- It has a user friendly design: the fine-tuning of the results can be carried out by changing the values of a few parameters only.

Although our Denton method is designed for the application to the accounting framework of Dutch Supply and Use table, it may also be of use in other application areas in which changes between time periods are considered more important than levels.

This chapter is organised as follows. In Section 2.2 the extended multivariate Denton model is presented. Section 2.3 describes how the model is applied at Statistics Netherlands. Section 2.4 concludes and gives an outlook on further research possibilities.

2.2 MODEL

We first present the univariate Denton method. Next, we describe the multivariate case and finally the extended multivariate Denton method is explained.

2.2.1 Univariate model

The aim of the classical Denton method is to find a benchmarked time series of scalars $\hat{x}_t$, $t = 1,\ldots,T$, that preserves as much as possible all quarter-to-quarter changes of the original quarterly time series x_t and that is subject to the annual benchmarks, y_a, $a = 1,\ldots,T/4$.

Denton proposed several measures to define the quarter-to-quarter changes. We consider the additive first-order difference function and the proportional first-order difference function. The additive function keeps additive differences $(\hat{x}_t - x_t)$ as constant as possible over all periods. The proportional function preserves the growth rates of x_t and therefore keeps the relative corrections $(\hat{x}_t - x_t)/x_t$ as constant as possible over all periods.

The objective function of the additive Denton model is

$$\min_{\hat{x}} \sum_{t=2}^{T} \left((\hat{x}_t - x_t) - (\hat{x}_{t-1} - x_{t-1}) \right)^2 \tag{2.1}$$

and the objective function of the proportional Denton model is

$$\min_{\hat{x}} \sum_{t=2}^{T} \left((\hat{x}_t / x_t) - (\hat{x}_{t-1} / x_{t-1}) \right)^2. \tag{2.2}$$

Both objective functions in (2.1) and (2.2) are subject to the following constraints

$$\sum_{t=4(a-1)+1}^{t=4(a-1)+4} \hat{x}_t = y_a, \qquad a = 1, \ldots, T/4 \tag{2.3}$$

where a is an index of the year and y_a is an annual value. The set of constraints expresses the alignment of four quarters to annual totals.

The proportional model cannot be used if the original time series contains zeroes. Although workarounds are possible, for instance replacing each zero by some very small number, it is strongly advised to use a different method for reconciling time-series with zeroes (e.g. an additive Denton method). Relative corrections are not defined in case of time-series with zeroes, and therefore it does not make sense to apply a criterion that is based on keeping those relative corrections as constant as possible.

Note that for the proportional model, the ratio $\hat{x}_t / x_t$ gives the relative change of a variable in time. When we approximate the original time series, we would like to preserve this change as much as possible. Therefore it would be more direct to consider the differences between the relative changes of the revised and preliminary series, i.e. to minimise the objective function $\sum_{t=2}^{T} \left((\hat{x}_t / \hat{x}_{t-1}) - (x_t / x_{t-1}) \right)^2$, which is the Causey-Trager growth rate preservation method (Bozik and Otto, 1988). However, this nonlinear form is very difficult to handle for large problems, see e.g. Öhlén (2006) and can be approximated with the function in (2.2). The reader is referred to Chapter 3 for a further discussion of the growth rate preservation method.

2.2.2 Multivariate case

In a multivariate setting a number of time series are benchmarked simultaneously. Again, quarterly figures are aligned with annual totals, but in addition there may also be constraints between related time series at each quarter.

A property of a multivariate model is that differences in reliability can be taken into account. This property is crucial for our application as the National Accounts use a wide range of sources. Naturally, different time series are considered more or less reliable, depending on their source. A multivariate model should adjust the quarterly changes of reliable time series less than those of unreliable ones. In the literature variances are often used in order to describe data reliability. Since in practice it is almost impossible to estimate these variances, we introduce weights instead. In our model weights can be viewed as generalisations of variances, i.e. they are defined in such a way that variances can substitute the weights. Analogous to variances, weights have to be strictly positive and satisfy the property that the higher the value, the more deviation is tolerated.

A multivariate model is formulated as follows. For an initial vector x_{it}, where x_{it} is the value of some time series i (i = 1,…,N), at some time period t (t = 1,…,T) that should satisfy a set of linear constraints, we aim to find a set of benchmarked time series $\hat{x}_{it}$ that satisfy all linear constraints, while preserving as much as possible all quarter-to-quarter changes of the original quarterly time series. The multivariate, additive model is given by

$$\min_{\hat{x}} \sum_{i=1}^{N} \sum_{t=2}^{T} \frac{1}{(w_{it}^{A})^2} \left((\hat{x}_{it} - x_{it}) - (\hat{x}_{it-1} - x_{it-1}) \right)^2, \tag{2.4}$$

$$\text{such that} \sum_{i=1}^{N} \sum_{t=2}^{T} c_{rit}^{H} \hat{x}_{it} = b_r^H, \qquad r = 1,\ldots,C^H \tag{2.5}$$

where w_{it}^{A} denotes a reliability weight of the i-th time series at quarter t and A stands for the additive model. How we define the weights will be described in detail in Subsection 2.2.4. In (2.5) r is the index of the constraints and C^H is the number of constraints. Further, c_{rit}^{H} are the coefficients of the constraints and b_r^H denote their target values. The superscript H stands for 'hard' constraints, we use it to distinguish these constraints from the 'soft' constraints that will be introduced in Subsection 2.2.3. Soft constraints must not be strictly adhered to, whereas for hard constraints violations are not acceptable.

The set of constraints in (2.5) may include two types of constraints: those that only affect data points within the same time step and those that span multiple time steps. The first type of constraints can be used to incorporate balancing constraints in the model. These relationships appear as a direct consequence of the economic accounting framework used. For instance, the National Accounts prescribe that total use and total supply have to be equal for each time period. The second type of constraints includes, amongst others, the annual alignment. For this type of constraints the annual values, y_a in (2.3), are included in b_r^H.

The univariate proportional model can be generalised to the multivariate case, similarly to the additive model. In Bikker and Buijtenhek (2006), the proportional and the additive models are combined, meaning that for each time series a choice for a proportional or an additive model has to be made. This choice has to be made beforehand and in practice it dependents on the content of the time series.

2.2.3 Extended model

In this subsection we define the extended multivariate Denton method. The extensions include: soft constraints, ratio constraints and inequalities. These constraints are added to the model, specifically for the application to the National Account data.

Based on knowledge and experience National Accounts specialists may have prior expectations with respect to the values some time series can attain. For instance, for perishable goods the value of the change of stocks, summed over the four quarters of one year, is expected to be close to zero. In order to include such knowledge in the model soft constraints are needed. A set of soft linear constraints is given by

$$\sum_{t=1}^{T}\sum_{i=1}^{N} c_{rit}^{S}\hat{x}_{it} \sim \left(b_r^S, w_r^L\right), \qquad r = 1,\ldots,L^S, \tag{2.6}$$

where L^S denotes the total number of linear constraints and b_r^S is a target value.

In the example of the perishable stocks b_r^S will be equal to zero and the summation is over four quarters. The superscript S stands for soft constraints and w_r^L is a reliability weight, where the superscript L indicates that the weight belongs to a linear constraint. Similar notation will be used throughout this section.

The constraints (2.6) are included in the model by adding the following penalization terms to the objective function in (2.4)

$$+\sum_{r=1}^{L^S} \frac{1}{\left(w_r^L\right)^2}\left(b_r^S - \sum_{t=1}^{T}\sum_{i=1}^{N} c_{rit}^{S}\hat{x}_{it}\right)^2. \tag{2.7}$$

Another important extension of the model is the ratio constraint. Many economic indicators are defined as ratios of National Account variables. For example, subject matter specialists may have prior expectations of the value of the ratio between value added and output of an industry. To describe these types of relations hard and soft ratio constraints are added to the model, that are given by

$$\begin{aligned} \hat{x}_{nt}/\hat{x}_{dt} &= v_{ndt} && n,d = 1,\ldots,N,\ t = 1,\ldots,T \\ \hat{x}_{nt}/\hat{x}_{dt} &\sim \left(v_{ndt},\ \left(w_{ndt}^R\right)^2\right) && n,d = 1,\ldots,N,\ t = 1,\ldots,T \end{aligned}$$

where w_{ndt}^R is the weight of a ratio of $\hat{x}_{nt}$ and $\hat{x}_{dt}$ and v_{ndt} is its predetermined target value.

Since we are unable to implement ratio constraints in their original form, we linearize these constraints first. Following the approach of Magnus *et al.* (2000), we obtain that

$$\begin{aligned} \hat{x}_{nt} - v_{ndt}\hat{x}_{dt} &= 0 \text{ and} \\ \hat{x}_{nt} - v_{ndt}\hat{x}_{dt} &\sim \left(v_{ndt},\ \left(w_{ndt}^{R*}\right)^2\right) \end{aligned} \tag{2.8}$$

where w_{ndt}^{R*} denotes the weight of a linearized ratio. The relation between w_{ndt}^{R} and w_{ndt}^{R*} is presented at the end of Subsection 2.2.4.

Hard linearized ratios are added to the constraints in the model, see (2.13) below. Soft linearized ratios are incorporated in the model, by adding the following term to the objective function

$$+\sum_{n,d=1}^{N}\sum_{t=1}^{T} B_{ndt}^{S}\left(\frac{\hat{x}_{nt}-v_{ndt}\hat{x}_{dt}}{w_{ndt}^{R*}}\right)^2, \tag{2.9}$$

where B_{ndt}^{S} is an indicator whose value is one if there is a soft ratio defined for $\hat{x}_{nt}$ and $\hat{x}_{dt}$ and zero otherwise.

Note that, essentially, there is no difference between linear constraints and linearized ratio constraints. The reason for making the distinction in the model, is that the weights will be defined in a different way. Contrary to the weights of linear constraints, the weights of the linearized ratio constraints depend on the target value v_{rt} (see Subsection 2.2.4).

Most economic variables cannot have negative signs. To incorporate these and other requirements in the model, inequality constraints are needed. A set of inequality constraints is given by

$$\sum_{i=1}^{N}\sum_{t=1}^{T} a_{rit}\hat{x}_{it} \leq z_r \qquad r = 1,\ldots, I^H, \tag{2.10}$$

where I^H denotes the number of inequality constraints and a_{rit} is a coefficient of $\hat{x}_{it}$.

Inequality constraints can easily be imposed in quadratic optimization problems, as it is proposed here for a multivariate Denton method. This extension is more complicated for other reconciliation methods. For instance, Boonstra *et al.* (2010) presented an approximation method of dealing with inequalities within the Bayesian macro integration method of Magnus *et al.* (2000), based on a truncated multivariate normal distribution. If we incorporate the terms defined in (2.7) and (2.9) in the objective function in (2.4), and add the constraints defined in (2.8) and (2.10) to (2.5), we obtain the complete, extended model

$$\min_{\hat{x}}\sum_{i=1}^{N}\sum_{t=2}^{T} A_{it}\left(\frac{(\hat{x}_{it}-x_{it})-(\hat{x}_{it-1}-x_{it-1})}{w_{it}^{A}}\right)^2+\sum_{i=1}^{N}\sum_{t=2}^{T}(1-A_{it})\left(\frac{1}{w_{it}^{P}}\left(\frac{\hat{x}_{it}}{x_{it}}-\frac{\hat{x}_{it-1}}{x_{it-1}}\right)\right)^2+$$
$$+\sum_{r=1}^{L^S}\frac{1}{\left(w_r^L\right)^2}\left(b_r^S-\sum_{t=1}^{T}\sum_{i=1}^{N}c_{rit}^{S}\hat{x}_{it}\right)^2+\sum_{n,d=1}^{N}\sum_{t=1}^{T}B_{ndt}^{S}\left(\frac{\hat{x}_{nt}-v_{rt}\hat{x}_{dt}}{w_{ndt}^{R*}}\right)^2 \tag{2.11}$$

such that

$$\sum_{i=1}^{N}\sum_{t=1}^{T} c_{rit}^{H}\hat{x}_{it} = b_r^H, \qquad r = 1,\ldots,C^H \tag{2.12}$$

$$B_{ndt}^{H}\left[(\hat{x}_{nt}-v_{ndt}\hat{x}_{dt}) = 0\right] \qquad n,d = 1,\ldots,N,\ t=1,\ldots,T \tag{2.13}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T} a_{rit}\hat{x}_{it} \leq z_r \qquad r = 1,\ldots, I^H, \tag{2.14}$$

Here A_{it} is an indicator function, defined as follows:

$$A_{it} = \begin{cases} 1 & \text{if the additive model is applied to series } i \\ 0 & \text{if the proportional model is applied to series } i \end{cases}$$

The four terms in the function (2.11) denote: additive quarterly changes, proportional changes, (soft) linear constraints and (soft) ratio constraints, respectively. The constraints in (2.12)–(2.14) denote: (hard) linear constraints, (hard) ratio constraints and inequality constraints, respectively.

The problem, defined by (2.11)–(2.14) is a standard convex quadratic programming (QP) problem. It is well known in the literature, and many efficient solving techniques are available (see e.g. Hillier and Lieberman, 2008 and Nocedal and Wright, 2006).

As in Bikker and Buijtenhek (2006), we determine beforehand which model, additive or proportional, is applied to each time series. Only a single model type can be assigned to a time series. For the National Account data the proportional model is preferred for most of the time series, as their data sources measure proportional growth rates. There are two exceptions:

1. If one of the quarterly values in absolute terms is less than some specified small value. Since in our application the preliminary time series are integer valued, when the initial values are small, relative changes may be heavily influenced by the preceding rounding process and therefore it does not make sense to preserve them. Another reason for this exception is that the proportional model cannot be used for time series that contain preliminary values of zero.
2. If a time series has both positive and negative values. When the proportional model is used, a result of the benchmarking could be that all values of a time series are multiplied by a negative number. Thus, all positive numbers become negative and vice versa. In practice this is not the desired outcome.

2.2.4 Weights

In the objective function of the aforementioned model several kind of weights are used (weights of additive and proportional changes, soft ratios and soft, linear constraints). In this subsection we define the weights used in (2.11) – (2.14).

Underlying model properties

We want our multivariate model to have the following properties:

Property 1) Invariance of input data (mentioned by Öhlén, 2006): If all input data are multiplied by the same nonnegative scalar, the outcomes must also be changed by this factor;

Property 2) Ratio symmetry (mentioned by Magnus and Danilov, 2008): The outcome does not change if a ratio in the benchmarking model is replaced by its reciprocal, i.e. if the constraint $x/y\sim\left(r,\left(w^R_{x/y}\right)^2\right)$ is replaced by $y/x\sim\left(1/r,\left(w^R_{y/x}\right)^2\right)$;

Property 3) Invariance of model choice: when the original time series is constant in time, i.e. if $x_{it} = x_i$ the results of the additive and proportional model should be the same;

The third property trivially holds true in the univariate case, but not in a multivariate setting. It prevents the results of the model to change more than necessary, when the model type is switched from additive to proportional or vice versa.

Proposed definitions

Below expressions are proposed for the weights of the quarterly mutations and the soft constraints. In Appendix A we show that these expressions satisfy the three above mentioned properties. Keeping in mind that the expressions should be easy to use, we introduce tuning parameters that apply to groups of similar weights.

Property 1 above implies that all terms in the objective function must be of the same dimension. While this still leaves an infinite number of choices for the weight expressions, the easiest choice is to use dimensionless (scalar) terms. For proportional quarter-to-quarter changes the squared weights are therefore defined by

$$\left(w_{it}^{P}\right)^2 = \left(\theta_i^{I}\right)^2, \tag{2.15}$$

where θ_i^I is a non-negative parameter that characterises the relative importance of time series x_i, compared to the other time series.

For the additive model the squared weights are defined by

$$\left(w_{it}^{A}\right)^2 = \left(\theta_i^{I}\right)^2 \bar{x_i^2}, \tag{2.16}$$

where

$$\bar{x_i^2} = \frac{1}{T}\sum_{t=1}^{T} x_{it}^2$$

is the mean squared value of x_i. Here, $\bar{x_i^2}$ is replaced by some value close to zero, if it is below some threshold value, since weights cannot be equal to zero. The expressions for the weights in (2.15) and (2.16) are chosen so, that the above mentioned Properties 1 and 3 are satisfied (see also Appendix 2.A).

The expression in (2.16) resembles the expression that is proposed by Beaulieu and Bartelsman (2006). The most important difference is that their definition involves x_{it}, where we use $\bar{x_i^2}$ instead. As a consequence the weights in (2.16) are the same for each time period t. The same applies to the weights in (2.15) for the proportional model. The motivation for this principle is that the reconciliation adjustments of one time series will be as constant as possible over time, just as in the original univariate method of Denton

(1971). We may also assume that the reliability of the source will not change in a short time period.

The parameter θ_i^I appears in all other weight expressions as well. By changing the value of it, all weights are adjusted that are related to the i-th time series, i.e. the weights of all quarterly changes of x_i, the weights of the linear constraints and the weights of the ratio constraints in which the time series x_i appears.

The expression of the squared weight of linear constraints is

$$\left(w_r^L\right)^2 = \left(\alpha^L \theta_r^L\right)^2 \frac{1}{\left(c_r^S\right)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(c_{rit}^S \theta_i^I x_{it}\right)^2, \tag{2.17}$$

where

$$c_r^S = \sum_{t=1}^{T} \sum_{i=1}^{N} \left(c_{rit}^S\right)^2, \tag{2.18}$$

In (2.17) α^L defines the "importance" of the model component linear constraints, in comparison with the other components (i.e. quarterly changes and ratio constraints). By decreasing the value of this parameter, all linear constraints will be made more important simultaneously.

Further, θ_r^L is a parameter that reflects the relative importance of one specific linear constraint, compared to the other linear constraints.

The last component in (2.17)

$$\frac{1}{\left(c_r^S\right)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(c_{rit}^S \theta_i^I x_{it}\right)^2, \tag{2.19}$$

is a weighted average of $\left(\theta_i^I x_{it}\right)^2$ The average in (2.19) is taken over all values of time series that appear in the constraint r. The weights are the squared coefficients c_{rit}^S of the constraint. So, the weight of a linear constraint (2.17) is determined by the average weights of all series in restriction r, corrected by a factor $\left(\alpha^L \theta_r^L\right)^2$.

We now continue with an expression for the weight of a ratio constraint. The weight of linearized ratio will be derived from an expression of the weight of a non-linearized ratio constraint.

The squared weight of a non-linearized ratio constraint will be assumed to be

$$\left(w_{ndt}^R\right)^2 = \left(\alpha^R \theta_{nd}^R\right)^2 \theta_n^I \theta_d^I \left(v_{ndt}\right)^2 \tag{2.20}$$

The components of this weight are quite similar to the components that are mentioned before. Similar to α^L in (2.17), the first factor α^R defines the relative "importance" of all ratio constraints, compared to the time series and the soft linear restrictions. The second factor θ_{nd}^R describes the relative importance of the specific ratio constraint, compared to the other ratio constraints. This factor is similar to θ_i^I in (2.17). The third and fourth component are the geometric mean of $\left(\theta_n^I\right)^2$ and $\left(\theta_d^I\right)^2$. It stands for the relative reliability of the time series that appear in the denominator and numerator of the ratio constraint. Finally, the fifth component v_{ndt}^2 denotes the square of the target value of the ratio.

The expression for the squared weight of a linearized ratio constraint (2.8) is

$$\left(w_{ndt}^{R*}\right)^2 = \left(\alpha^R \theta_{nd}^R\right)^2 \theta_n^I \theta_d^I (v_{ndt} \tilde{x}_{dt})^2, \tag{2.21}$$

where $\tilde{x}_{dt}$ is defined below in (2.24). The expression (2.21) has been derived from (2.20). It follows from

$$\frac{\frac{\hat{x}_{nt}}{\hat{x}_{dt}} - v_{ndt}}{w_{ndt}^R} = \frac{\hat{x}_{nt} - v_{ndt}\hat{x}_{dt}}{w_{ndt}^R \hat{x}_{dt}} \tag{2.22}$$

In the left hand side of (2.22) stands the root of one term of the objective function, corresponding to a non-linearized ratio. The numerator of the right hand side of (2.22) is a linearized ratio as it appears in the objective function (2.9). By definition, the denominator of the right hand side is the weight of the linearized ratio. Thus, it follows that

$$w_{ndt}^{R^*} = w_{ndt}^R \hat{x}_{dt}. \tag{2.23}$$

The expression (2.23) cannot be used in practice, because $\hat{x}_{dt}$ denotes a variable, whose value is not known prior to the benchmarking. Therefore we replace $\hat{x}_{dt}$ in (2.23) by

$$\tilde{x}_{dt} = \frac{1}{1 + (v_{ndt})^2} x_{dt} + \frac{(v_{ndt})^2}{1 + (v_{ndt})^2} \frac{x_{nt}}{v_{ndt}}, \tag{2.24}$$

a weighted average of x_{dt} and x_{nt}/v_{ndt}, as proposed by Magnus and Danilov (2008).

2.2.5 Example

In order to demonstrate how we define the optimization function in practice, let us consider a benchmarking problem, consisting of 12 quarters of two time series x_1 and x_2. Suppose that each quarterly value is equal to 10 and that annual benchmarks are available for both time series. The annual figures are 50, 75 and 95 for the three consecutive years. These annual figures are the same for both time series. This example is not very realistic, we intentionally choose an example with large discrepancies between the quarterly and annual data in order to illustrate more vividly how the benchmarking model works. Now assume that the first annual alignment is binding, whereas the second and the third are not. Furthermore we know that the figures from the first series should approximately be 10% higher than the figures from the second time series.

Now we can formulate the optimization model. The constraints for the first year are binding, which means that we have the hard constraints

$$\sum_{t=1}^{4} \hat{x}_{it} = 50, \quad i = 1,2.$$

For the second and the third year the constraints are not binding, hence we have the following soft constraints

$$\sum_{t=5}^{8} \hat{x}_{it} \approx 75 \text{ and } \sum_{t=9}^{12} \hat{x}_{it} \approx 95 \quad i = 1,2.$$

Furthermore, there is one soft ratio constraint, defined by

$$\frac{\hat{x}_{1t}}{\hat{x}_{2t}} \approx 1.1 \quad t = 1,\ldots,12$$

and we use the proportional model for both time series. Note that the soft, ratio constraint is inconsistent with the annual figures of both time series. The relative values of the weights of these model components will determine which model component influences the outcome most. The optimization model can be written as follows

$$\begin{aligned}
&\min_{\hat{x}} \sum_{t=2}^{12} \left(\frac{1}{w_{1t}^{P}} \left(\frac{\hat{x}_{1t}}{10} - \frac{\hat{x}_{1t-1}}{10} \right) \right)^2 + \sum_{t=2}^{12} \left(\frac{1}{w_{2t}^{P}} \left(\frac{\hat{x}_{2t}}{10} - \frac{\hat{x}_{2t-1}}{10} \right) \right)^2 \\
&+ \left(\frac{75 - \Sigma_{t=5}^{8} \hat{x}_{1t}}{w_1^L} \right)^2 + \left(\frac{75 - \Sigma_{t=5}^{8} \hat{x}_{2t}}{w_2^L} \right)^2 + \left(\frac{95 - \Sigma_{t=9}^{12} \hat{x}_{1t}}{w_3^L} \right)^2 \\
&+ \left(\frac{95 - \Sigma_{t=9}^{12} \hat{x}_{2t}}{w_4^L} \right)^2 + \sum_{t=1}^{12} \left(\frac{\hat{x}_{1t} - 1.1\,\hat{x}_{2t}}{w_{12t}^{R*}} \right)^2 \\
&\text{such that} \\
&\sum_{t=1}^{4} \hat{x}_{1t} = 50 \text{ and} \\
&\sum_{t=1}^{4} \hat{x}_{2t} = 50.
\end{aligned}$$

The parameters of the weights and their values are given in Table 2.1. For simplicity we choose θ_1^I, θ_2^I and α^R equal to one.

Since parameters θ_1^I and θ_2^I have the same value, both time series are considered equally reliable. Their parameter values imply that $(w_{it}^P)^2 = 1$ for all i and t. Since the

Table 2.1. Weight parameters

	Series		Annual alignment		Ratio	
	θ_1^I	θ_2^I	α^L	θ_1^L	α^R	θ_{12}^R
Value	1	1	2	0.5	1	0.5

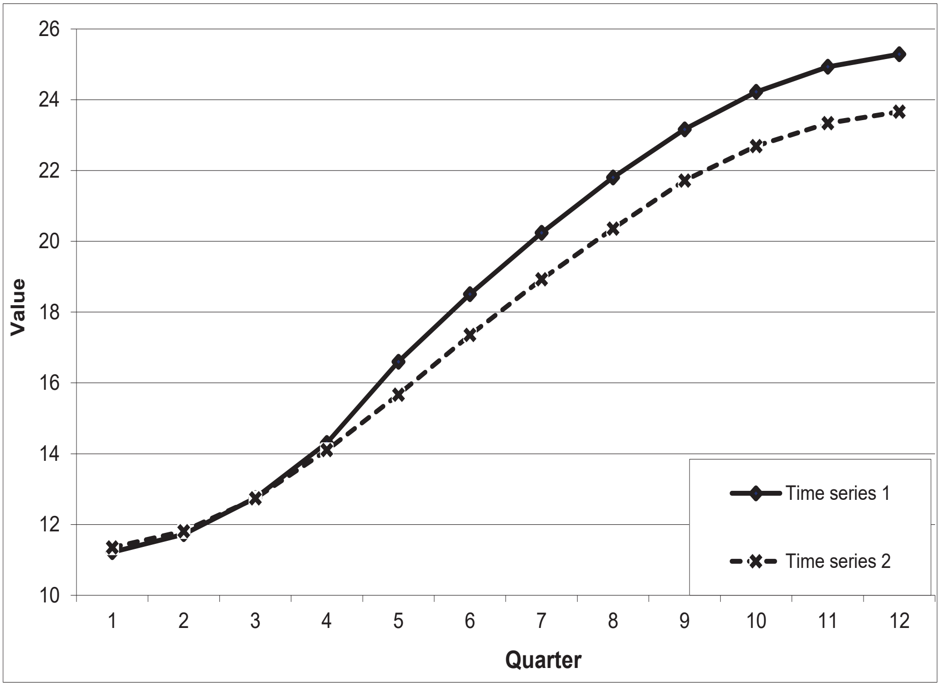

Figure 2.1 Benchmarked series

average value of $\left(\theta_i^I x_{it}\right)^2$ over the four quarters of one year is 100 for both time series, it follows that $\left(w_r^L\right)^2 = 100$ for all r. The ratio parameters imply $\left(w_{12t}^R\right)^2 = 0.30$ for all t. Since $\bar{x}_{2t} = 9.502$ for all t, it follows that $\left(w_{12t}^{R*}\right)^2 = 27.31$.

The results of the benchmarking method, depicted in Figure 2.1, are two time series, whose values gradually increase over time. This increase is due to the annual benchmarks. Further note that as a result of the ratio constraint, $\hat{x}_{1t}$ increases more rapidly than $\hat{x}_{2t}$ from the fifth quarter onwards. During the first four quarters, the influence of the ratio constraint is negligible, since the quarters of both time series have to strictly add up to the same annual values. In the second and third year the annual alignment is soft, and therefore the ratio constraint is more important than for the first year.

Table 2.2 shows that the reconciled annual figures of the second and third year closely approximate their target values.

Suppose we decrease θ_1^I from 1 to 1/2, and θ_2^I is left untouched. As a consequence, time series 1 becomes more "important" compared to time series 2, amongst others, the annual alignment of time series 1 becomes more tight.

Table 2.3 indeed shows that the reconciled, annual figures of time series 1 approximate their target values more closely, compared to Table 2.2, while the opposite holds true for the second time series.

Table 2.2 Annual values (benchmarked); $\theta_1^I = 1.$

	Year 1	Year 2	Year 3
Time Series 1	50.00	77.16	97.61
Time Series 2	50.00	72.32	91.42
Target value (both series)	50.00	75.00	95.00
Time Series 1 / Time Series 2	1.000	1.067	1.068
Target value (ratio)	1.100	1.100	1.100

Table 2.3 Annual values (benchmarked), $\theta_1^I = 1/2$

	Year 1	Year 2	Year 3
Time Series 1	50.00	75.81	95.88
Time Series 2	50.00	70.72	89.37
Target value (both series)	50.00	75.00	95.00
Time Series 1 / Time Series 2	1.000	1.072	1.073
Target value (ratio)	1.100	1.100	1.100

Furthermore, the ratio constraint becomes somewhat more important, compared to the case of the initial parameters values, since the weights of the ratio constraints are positively correlated to the average value of θ_1^I and θ_2^I. This can be seen by comparing Table 2.3 with Table 2.2. The benchmarked value of the ratio approximates its target values of 1.1, more closely in Table 2.3.

2.3 APPLICATION

The model is very well suited for application to real life statistical data, yet one has to bear in mind that the basic assumption under any least-squares model is that the statistical discrepancies are independently distributed with a mean of zero. However, large discrepancies are usually not caused by sampling errors. They are not independently distributed with mean zero and therefore cannot be reconciled by a least squares method.

We therefore apply the model in a two-step process: first we detect and correct large discrepancies, then we apply benchmarking for smoothing out the remaining differences. In the first step subject matter specialists solve the large discrepancies. To achieve this, they can use several sources of information, like earlier estimates or information on how the data sources were compiled. The remaining smaller discrepancies may still have arisen both from errors and sampling noise. Yet, when all discrepancies are small, this distinction is practically irrelevant. The results of benchmarking will generally be acceptable irrespective of the cause of the discrepancies. By being able to focus on the large problems only, time and effort is saved. The exact definition of 'large' is a trade-off between quality and cost.

In order to be useful for practical implementation at Statistics Netherlands, the benchmarking software has to be able to cope with very large data sets. Statistics Netherlands has built benchmarking software, using XPRESS (FICO, 2009) as a solver. This state-of-the-art, commercial optimization solver is able to cope with very large data sets. A model based on the Dutch supply and use tables with 51 832 time series, each consisting of up to 3 annual, and 12 quarterly values, was translated into a quadratic optimization problem with 503 451 free variables and 163 792 constraints. By using XPRESS on a PC with 2.0 GHZ, Xeon E5335 with 2048 MB Ram, the optimal solution was found in approximately one hour and a half. The capacity of the benchmarking software is further limited by the computer memory available.

The benchmarking software solves a general quadratic optimization model, which is an abstraction of the statistical benchmarking model. In order to be able to specify the optimization model in economic and statistical terms, we implemented a separate software module. This module consists of a software library which can be incorporated in any scripting programming language. The library offers a data model which basically consists of a collection of time series and a collection of constraints. The user reads the time series from data files and specifies the constraints using routines in a script. To ease the specification of constraints, the library offers methods for searching and grouping time series, based on their classifications or names. The library also helps specifying the many parameters of the model, for instance the reliability weights of time series, non-binding annuals and soft constraints. Thanks to this, changing the values of a few parameters is usually sufficient for fine-tuning the results. The structure of the accounting framework does not change much from year to year, so, once created, a script can be re-used with slight modifications for many years.

This flexible way of specifying the optimization model makes it possible to incorporate a wide range of statistical and economic relationships. The same software can therefore be used for several types of National Accounts or even be applied outside of the National Accounts. By combining model elements like hard and soft linear constraints, inequality constraints, ratios, the additive and proportional model type and reliability weights, we can make elaborate modelling constructions. Ratios can be used for the relation between current and constant price time series and also for structural relations, like those between the volume growth of taxes on goods and goods themselves. It is also possible to define new time series in terms of existing ones in the script. Being able to use these derived time series in constraints greatly enhances the modelling possibilities.

Annual totals can be missing in the benchmarking model. Generally, only one annual total per time series suffices for obtaining a solution. In this case a constraint for the missing annual total will be simply omitted from the optimization model. We use this property of the model to estimate recent annual supply-use tables based on incomplete new annual information and earlier quarterly estimates. The model also allows for time series of different length. In our standard setup for the supply-use tables, the total benchmarking period consists of three years, yet constant price time series by definition exist for two years only. In our implementation, time series are constructed by piecing together partial series consisting of only four quarterly values and (optionally) an annual total. The quarterly changes between two consecutive years may or may not be preserved. Therefore the user is free to specify the length of the series.

With our new software, the task of implementing the economic and statistical rules lies with statisticians, not with the programmers. In our experience this is both a trial and a blessing. The initial effort to build and test a complete setup is particularly costly in terms of time. However, once a working model setup is implemented in a script, making changes or extensions is quite easy. Even a complete change in the classifications of the variables is relatively easy to incorporate.

Special consideration is given to checking the outcomes. A mistake in the rules can easily remain hidden in the outcomes, due to the sheer bulk of the datasets. We therefore implemented several automated and manual checks, both during the process and at the end. For instance, the benchmarking software generates tables with scores that show when quarter-to-quarter changes in individual time series are adjusted too much, or when the outcomes cannot be made to fit soft constraints very well. The script that builds the optimization problem can also be used for problem detection. For instance, it can give warnings when it detects large discrepancies in the data or when inconsistent constraints are specified. It can also be used for automatic documentation of the applied rules, so a statistician can check them. Thus, the role of the statistician changes from doing the actual data reconciliation to specifying the input of the model and checking the results.

2.4 CONCLUSIONS

Statistical agencies publish both annual and quarterly figures. Achieving consistency between these can be highly labour intensive job, especially when the figures are part of an accounting framework and must adhere to accounting rules. In this chapter we present a model which achieves this consistency with minimal adjustment to the data. The model is the generalised multivariate Denton model. In this model we brought together different building blocks like linear constraints, inequality constraints, ratios and reliability weights. By combining these elements we can make elaborate modelling constructions, thus creating a very flexible and powerful benchmarking instrument.

For an application to the supply and use tables of Statistics Netherlands a problem has to be solved, consisting of over 500 thousand free figures. Examples of benchmarking of such large-scale applications in the literature are rare. The generalised multivariate Denton method was successfully implemented in the production process at Statistics Netherlands in 2010. The introduction of this method drastically changed the reconciliation process. The role of the statistician now is to check the input data, the model setup and the results. Using this benchmarking instrument for the labour intensive tasks has freed up time which can be used for checking and improving the quality of the results.

APPENDIX 2.A PROOF OF THE PROPERTIES OF THE MODEL

In Subsection 2.2.4 we presented three desired properties of a benchmarking model. Here, we show that the model indeed satisfies the properties: 1) invariance of input data, 2) symmetry of ratios and 3) invariance of model choice.

Invariance of input data

Invariance of input data means that the multiplication of all input data by the same nonnegative scalar, leads to outcomes that are changed by the same factor. The model we propose satisfies this property, since multiplying each of the variables x_{it}, $\hat{x}_{it}$, b_r^H, b_r^S and z by a nonnegative scalar λ does not change the objective function and the constraints of the model. For instance, for the part of the objective function that describes the additive mutations, it holds true that

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{\left((\lambda\hat{x}_{it}-\lambda\hat{x}_{it-1})-(\lambda x_{it}-\lambda x_{it-1})\right)^2}{\left(\theta_i^A\right)^2\frac{1}{T}\sum_{u=1}^{T}(\lambda x_{iu})^2}$$
$$=\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{\left((\hat{x}_{it}-\hat{x}_{it-1})-(x_{it}-x_{it-1})\right)^2}{\left(\theta_i^A\right)^2\frac{1}{T}\sum_{u=1}^{T}(x_{iu})^2}.$$

It is easy to show that the other parts of the model also satisfy this property.

Symmetry of ratios

Symmetry of ratios means that it does not matter for the results whether a soft ratio constraint is defined by

$$\hat{x}_n/\hat{x}_d \approx v,$$

or by its reciprocal

$$\hat{x}_d/\hat{x}_n \approx 1/v,$$

For convenience, some of the subscripts of $\hat{x}_n$, $\hat{x}_d$ and v are omitted. The corresponding terms in the objective function are

$$\left(\frac{\hat{x}_n - v\hat{x}_d}{\rho v\hat{x}_d}\right)^2 \tag{2.25}$$

and

$$\left(\frac{\hat{x}_d - \frac{1}{v}\hat{x}_n}{\rho\frac{1}{v}\hat{x}_n}\right)^2 \tag{2.26}$$

where $\rho^2 = \left(\alpha^R \theta^R_{nd} \theta^I_n \theta^I_d\right)^2$. The property of symmetry is satisfied if (2.25) and (2.26) are the same. Below it is shown that this holds true. Analogous to the definition (2.24) of $\breve{x}_d$ in (2.25), the definition of $\breve{x}_n$ in (2.26) is

$$\breve{x}_n = \frac{1}{1+\frac{1}{v^2}} x_n + \frac{\frac{1}{v^2}}{1+\frac{1}{v^2}} v x_d. \tag{2.27}$$

The definition (2.27) is obtained from (2.24) by interchanging x_n and x_d and replacing v by $1/v$. Note that

$$\breve{x}_n = \frac{v^2}{v^2+1} x_n + \frac{1}{v^2+1} v x_d = v \breve{x}_d,$$

which follows from the definition in (2.24). By using this result, it follows that

$$\left(\frac{\hat{x}_d - \frac{1}{v}\hat{x}_n}{\rho \frac{1}{v}\hat{x}_n}\right)^2 = \left(\frac{\hat{x}_d - \frac{1}{v}\hat{x}_n}{\rho \hat{x}_d}\right)^2 = \left(\frac{\hat{x}_n - v\hat{x}_d}{\rho v \hat{x}_d}\right)^2,$$

or equivalently that (2.25) is the same as (2.26), which proofs the symmetry of ratio property.

Invariance of model choice for constant time series

The invariance of model choice property means that the results of the additive model and the proportional model are the same for constant time series. Consider some time series i and suppose that its initial quarterly values are constant, i.e. $x_{it} = x_i$ for all t, then one component of the objective function of the additive model can be rewritten by

$$\frac{\left((\hat{x}_{it} - \hat{x}_{it\text{-}1}) - (x_{it} - x_{it\text{-}1})\right)^2}{w^A_{it}} = \left(\left(\frac{\hat{x}_{it}}{x_{it}} - \frac{\hat{x}_{it\text{-}1}}{x_{it\text{-}1}}\right) / \left(\frac{w^A_{it}}{x_{it}}\right)\right)^2, \tag{2.28}$$

and the corresponding component of the objective function of the proportional model is

$$\left(\left(\frac{\hat{x}_{it}}{x_{it}} - \frac{\hat{x}_{it\text{-}1}}{x_{it\text{-}1}}\right) / w^P_{it}\right)^2, \tag{2.29}$$

The invariance of model choice property is fulfilled, if (2.28) and (2.29) are the same. That is, if

$$\left(w^P_{it}\right)^2 = \left(w^A_{it} / x_{it}\right)^2. \tag{2.30}$$

Under the assumption $x_{it} = x_i$ for all t, the definition of $\left(w^A_{it}\right)^2$ in (2.16) can be written as

$$\left(w^A_{it}\right)^2 = \left(\theta^I_i\right)^2 \frac{1}{T} \sum_{u=1}^{T} x^2_{iu} = \left(\theta^I_i\right)^2 \frac{1}{T}\left(T \bar{x}^2_{it}\right) = \left(\theta^I_i\right)^2 \bar{x}^2_{it} \tag{2.31}$$

and by combining this with the definition of $(w_{it}^{p})^2$ in (2.15) we obtain that our model indeed satisfies the invariance of model choice property.

APPENDIX 2.B PRACTICAL APPLICATION

The parameters θ_i^I in the various expressions of the weights must be given a value. When available, weights can be based on reliability measures such as variances. In practice however, this information is usually missing. In this situation it is still possible to apply the model, using weights based on a subjective estimate of the relative reliability.

To achieve this, collections of series must be classified in discrete classes of subjective reliability, ranging from “Least reliable” to “Most reliable”. We can then map the classes onto values for θ_r^L and θ_{ndt}^R, using the following expressions:

$$\begin{aligned}
\left(\theta_i^I\right)^2 &= \beta^{-2J_i}, \\
\left(\theta_r^L\right)^2 &= \beta^{-2L_r}, \\
\left(\theta_{ndt}^R\right)^2 &= \beta^{-2R_{ndt}},
\end{aligned}$$

where $\beta > 1$ and $J_i,\ L_r,\ R_{ndt} \in \mathbb{Z}$ with $|J_i|,\ |L_r|,\ |R_{ndt}| \leq K$. These definitions guarantee that all $\theta_i^I\ \theta_r^L$ and θ_{ndt}^R are nonnegative.

The meaning of the parameters is as follows: β determines the degree of variation of the weights, J_i describes the relative reliabilities of the different time series, L_r describes the relative importance of the specific constraint r compared to the other linear constraints, and R_{ndt} is the equivalent parameter for ratios.

Finding the most appropriate values of these tuning parameters is a trial and error process that depends on the desired outcome of the model, which may be different from application to application. An important aspect in choosing the values for β and J_i is the scaling of the variables in the problem. A computer cannot represent infinitely small differences between numbers. The smallest difference is about $1*10^{-16}$. This limit implies certain bounds on β and J_i. For time series with values in six digits the term $\left(w_{it}^P x_{it}\right)^2$ in the objective function (2.11) may have more than 12 digits (depending on the value of w_{it}^P), while other time series may have small values and a small weight, leading to a term in the objective function that has 2-4 digits behind the decimal point. The values for β and J_i should be chosen so that the number of significant digits in the values in the objective function in (2.11) does not exceed the maximum capacity of 16 digits. In particular this means that the value of β should not be too large. A value for β can easily be determined in a simulation experiment. In our application we typically use β between 1.5 and 2.0.

The values of for the parameters J_i, L_r and R_{nd} of relative reliability are determined by subject matter specialist. In our applications we use $K = 3$, meaning that J_i, L_r, $R_{nd} \in \{-3, -2, -1, 0, 1, 2, 3\}$. For the time series obtained from the most reliable sources $J_i = 3$ meaning that these time series have the smallest weight. The choice $K = 3$, is made for ease of use: seven classes of relative reliability turned out to be manageable in practice. The values of the parameters α^R and α^L are 1 in our applications. This is the default value, we did not see the necessity of adjusting this value.

3. GRP temporal benchmarking: drawbacks and alternative solutions[2]

Summary. Benchmarking monthly or quarterly series to annual data is a common practice in many National Statistical Institutes. The benchmarking problem arises when time series data for the same target variable are measured at different frequencies and there is a need to remove discrepancies between the sums of the sub-annual values and their annual benchmarks. Several benchmarking methods are available in the literature. The Growth Rates Preservation (GRP) benchmarking procedure is often considered the best method. It is often claimed that this procedure is grounded on an ideal movement preservation principle. However, we show that there are important drawbacks to GRP, relevant for practical applications, that are unknown in the literature. Alternative benchmarking models will be considered that do not suffer from some of GRP's side effects.

2 This chapter has been published as: Daalmans J.A, T. di Fonzo, N. Mushkudiani and R.P. Bikker (2018) GRP temporal benchmarking: drawbacks and alternative solutions. *Survey Methodology,* 44, 43-60.

3.1 INTRODUCTION

Benchmarking monthly and quarterly series to annual data is a common practice in many National Statistical Institutes. For example, each year Statistics Netherlands aligns 12 quarterly Supply and Use Tables with the three most recent annual accounts (Eurostat, 2013, Annex 8C).

The benchmarking problem arises when time series data for the same target variable are measured at different frequencies with different levels of accuracy. One might expect that a temporal aggregation relationship between these time series is fulfilled, e.g. that four quarterly values add up to one annual value, but because of differences in data sources and processing methods, this is often not the case. Benchmarking is the process to remove such discrepancies. In this process the preliminary values are adjusted to achieve mathematical consistency between low-frequency (e.g., annual) and high-frequency (e.g., quarterly or monthly) time series.

There are two main principles of benchmarking. Firstly, low-frequency benchmarks are fixed, because these data sources describe levels and long-term trends better than high-frequency sources. Secondly, short-term movements of high-frequency time series are preserved as much as possible, as these data sources provide the only information on short-term movements.

Several benchmarking methods are available in the literature. These methods differ in the way short-term movements of high-frequency series are defined. A distinction can be made between multiplicative and additive methods. Multiplicative methods try to preserve relative changes of preliminary high-frequency time series, while additive methods aim to preserve changes in absolute terms. In this chapter the focus will be solely on multiplicative variants.

Two well-known multiplicative methods are Denton Proportionate First Differences (PFD), by Denton (1971), and Growth Rates Preservation (GRP) by Causey and Trager (1981; see also Trager, 1982, and Bozik and Otto, 1988).

In the literature it is generally agreed that GRP is grounded on the strongest theoretical foundation (Bloem *et al*. 2001, p 100). It explicitly preserves the period-to-period rates of change of the preliminary series. However, Denton PFD is more popularly used, because it is technically easier to apply. Mathematically, the Denton method deals with a standard linearly constrained quadratic optimization problem, while GRP solves a more difficult linearly constrained nonlinear problem that can be efficiently solved by an interior-point-algorithm (Di Fonzo and Marini, 2015).

From a number of simulation studies it is known that Denton PFD and GRP lead to similar or close to similar results for the large majority of cases (Harvill Hood, 2005, Titova *et al.*, 2010, Di Fonzo and Marini, 2012a, and Daalmans and Di Fonzo, 2014). Therefore Denton PFD can be used as an approximation of GRP.

The aim of this chapter is to demonstrate that GRP suffers from drawbacks that are, to the best of our knowledge, not described in the literature. A first drawback is that it matters whether benchmarking is applied ‘forward’ or ‘backward’ in time. In this context, we will present a link with the time reversibility property from index number theory. A second drawback is that undesirable results may be obtained due to singularities in the GRP objective function.

A second aim of this chapter is to present alternative benchmarking methods that do satisfy time reversibility. This chapter may be valuable for practitioners who apply or consider to apply benchmarking techniques.

First, Section 3.2 gives a formal description of the Denton PFD and GRP benchmarking methods. Section 3.3 describes the drawbacks of the GRP method. In Section 3.4 two new benchmarking methods are proposed that can be used as an alternative for GRP. Results of an illustrative application to real-life data are given in Section 3.5. Finally, Section 3.6 concludes this chapter.

3.2 TEMPORAL BENCHMARKING METHODS

This section explains the Denton PFD and GRP benchmarking procedures. Because temporal aggregation constraints are the same for Denton PFD and GRP, these are described first. Thereafter, the Denton PFD and GRP benchmarking procedures are explained.

We focus on univariate variants of these methods, in which temporal consistency is the main constraint of interest. The observations that are presented in the remainder of this chapter are however also valid for the multivariate case, in which multiple time-series are reconciled simultaneously and additional constraints between time-series apply (see Di Fonzo and Marini 2011 and Bikker *et al.* 2013).

3.2.1 General notation and temporal constraints

In general, temporal aggregation constraints can be expressed as a linear system of equalities $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{x}$ is the target vector of high-frequency values, $\boldsymbol{b}$ is a vector of low-frequency values and $\mathbf{A}$ is a temporal aggregation matrix converting high- into low-frequency values.

The specific form of these constraints depends on the nature of the variables involved. For flow variables, a sum of subannual values, e.g. four quarterly values, usually needs to be the same as one annual value. For stock variables, one of the subannual values, usually the first or the last, needs to be the same as the relevant annual value. For example, for quarterly/annual flow variables, assuming for the sake of simplicity that the available time span begins on the first quarter of the first year and ends on the fourth quarter of the last observed year, it is

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 & 1 \end{bmatrix}$$

Denoting by $\boldsymbol{p}$ a vector of preliminary values, in general it is $\mathbf{A}\boldsymbol{p} \neq \boldsymbol{b}$, otherwise no adjustment would be needed. We look for a vector of benchmarked estimates $\boldsymbol{x}^*$, a particular outcome for $\boldsymbol{x}$, which should be `as close as possible' to the preliminary values and that satisfies $\mathbf{A}\boldsymbol{x}^* = \boldsymbol{b}$.

3

Not all sub annual periods need to be covered by a benchmark. Thus, the number of rows in **A** may be smaller than the total number of annual periods, see e.g. Dagum and Cholette (2006) for more details.

In a benchmarking operation, characteristics of the original series $\boldsymbol{p}$ should be considered. For example, in an economic time series framework, the preservation of the temporal dynamics (however defined) of the preliminary series is often a major interest of the practitioner.

3.2.2 Growth Rates Preservation (GRP) and Denton PFD

This section gives a formal description of GRP and Denton PFD.

Causey and Trager (1981; see also Monsour and Trager, 1979, and Trager, 1982) obtain the benchmarked values x_t^*, t=1,…,n as a solution to the following optimization problem:

$$\min_{\boldsymbol{x}} f_F^{GRP}(\boldsymbol{x}) \text{ subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b},$$
where
$$f_F^{GRP}(\mathbf{x}) = \sum_{t=2}^{n}\left(\frac{x_t}{x_{t-1}} - \frac{p_t}{p_{t-1}}\right)^2. \tag{3.1}$$

The GRP criterion to be minimized, $f_F^{GRP}(\boldsymbol{x})$, explicitly relates to growth rates: it minimizes the sum of squared differences between growth rates of preliminary and benchmarked values. The subscript F in the minimization function stands for "Forward", later in this chapter a "Backward" minimization function will be defined.

Denton (1971) proposed a benchmarking procedure grounded on the Proportionate First Differences (PFD) between target and original series. Cholette (1984) slightly modified the result of Denton, in order to correctly deal with the starting conditions of the problem. The PFD benchmarked estimates are thus obtained as the solution to the constrained quadratic minimization problem

$$\min_{x_t} f_F^{PFD}(\boldsymbol{x}) \text{ subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b},$$
where
$$f_F^{PFD}(\boldsymbol{x}) = \sum_{t=2}^{n}\left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)^2. \tag{3.2}$$

The Denton PFD criterion to be minimized, $f_F^{PFD}(\boldsymbol{x})$, is a sum of squared linear terms, which is easier to deal with than the nonlinear GRP objective function.

3.3 TWO PROBLEMS WITH GRP BENCHMARKING

3.3.1 Time reversibility

Time reversibility means that it does not matter whether a method is applied forward or backward in time. This property can be of interest in many application areas.

In physics, it means that if time would run backwards, all motions are reversed. In index number theory, time reversibility was introduced in a classical work of Fisher (1922, page 64). It is stated that "if taking 1913 as a base and going forward to 1918, we find that, on the average, prices have doubled, then, by proceeding in the reverse direction, we ought to find the 1913 price level to be half that of 1918". The motivation of this principle is that the direction of time can be considered arbitrary; it does not have any naturally preferred direction.

Time reversibility can also be applied in the context of benchmarking. It means that if we would reverse a time series, apply benchmarking, and reverse the benchmarked series back again, we get exactly the same results as for benchmarking the original series. In other words: from the benchmarked results it cannot be seen whether benchmarking has been applied forward or backward in time.

Benchmarking a reversed time series, according to GRP and Denton PFD, respectively, is equivalent to minimizing the following objective functions

$$f_B^{GRP}(\boldsymbol{x}) = \sum_{t=2}^{n}\left(\frac{x_{t-1}}{x_t} - \frac{p_{t-1}}{p_t}\right)^2 \tag{3.3}$$

and

$$f_B^{PFD}(\boldsymbol{x}) = \sum_{t=2}^{n}\left(\frac{x_{t-1}}{p_{t-1}} - \frac{x_t}{p_t}\right)^2, \tag{3.4}$$

where subscript B stands for backwards. These objective functions are obtained from the forward objective functions by interchanging t and $t-1$. From now on, the minimization of (3.3) or (3.4) will be called 'backward benchmarking', as opposed to standard, forward benchmarking.

As mentioned above, a benchmarking method satisfies the time reversibility property if forward and backward benchmarking lead to the same results. It can be easily seen that $f_F^{GRP}(\boldsymbol{x}) \neq f_B^{GRP}(\boldsymbol{x})$, while $f_F^{PFD}(\boldsymbol{x}) = f_B^{PFD}(\boldsymbol{x})$. From this it follows that Denton PFD satisfies the time reversibility property, but GRP does not.

More practically, in many production processes "forward" benchmarking is applied, for example for the reconciliation of the Dutch Supply and Use tables (Bikker *et al.*, 2013). However, after a revision, revised time series may be constructed 'back in time', by using backward objective functions. To prevent ambiguity, it is highly undesirable that there are any differences in outcomes that can be purely attributed to a difference in 'time direction'. Practitioners who are unaware of the time reversibility property, may apply forward and backward benchmarking and mistakenly assume that both methods lead to the same results.

Although it is true that any benchmarking application can be restricted to preserving forward growth rates, it is undesirable that results are affected by the irrelevant property of time direction. Therefore, any benchmarking method should preferably satisfy time reversibility. Moreover, Subsection 3.3.3 illustrates that a benchmarking method that is not symmetric in time may change the timing of the most important economic events, e.g. the peaks and troughs that demark the start and end of a crisis.

3.3.1 Singularity

A second problem of GRP is the singularity of its objective function. If x_{t-1} approaches to zero in case of forward benchmarking (or x_t for backward benchmarking) the objective function value tends to infinity. This causes several problems.

One complication is that the optimization problem becomes unstable, a small change in preliminary values can lead to a large shift in benchmarked values. Consequently, undesirably large revisions can be obtained when benchmarking updated data.

Another complication is that, since a correction of near zero values can be heavily penalised, growth rates of such values are strongly preserved. This may however come at the expense of relatively large corrections of other growth rates. On the other hand, one may argue that growth rates do not contain much information for extremely small (close-to-zero) values. Hence, growth rate preservation can be deemed inappropriate in this case. Subsection 3.5.3 shows a real-life example of this problem.

A third complication is that, as close-to-zero benchmarked values may cause a large objective function value, GRP methods tend to avoid such values. Consequently, irregular correction patterns can be obtained. In particular, negative benchmarked values may be obtained for a problem in which all preliminary values are positive. Consider an example in which two consecutive values are 100. Then, an adjustment of the first value from 100 to –100 is less costly in terms of GRP's objective function value than a correction from 100 to 30. The corresponding objective function values are $((100/-100) - (100/100))^2 = 4$ and $((100/30) - (100/100))^2 = 5.44$. A value that goes from a large positive to a large negative will however usually not be considered good movement preservation. Therefore, the example also demonstrates the questionability of the use of growth rates when positive and negative values occur.

For this reason, it can be advisable to avoid negative outcomes by inclusion of non-negativity constraints, see Subsection 3.4.1 for more details. For Denton PFD negative values are less likely obtained. In the previous example, an adjustment from 100 to 30 is preferred to an adjustment from 100 to -100. A real-life example of this problem is shown in Subsection 3.5.3.

Although singularity of GRP's objective function may trigger negative benchmarked values, it is not the only cause. Denton PFD may also yield negative values. In general, there is a risk of negative benchmarked values, when the (relative) change from one benchmark to another significantly differs from the (relative) change from the underlying annualised preliminary values.

A fourth complication of GRP´s singular objective function is that irregular peaks and throughs may occur in a benchmarked time series. The explanation is that in standard GRP a correction of large positive value to a close-to-zero value is less costly in terms of the objective function value than an opposite correction from close-to-zero to a large positive. That is, a correction of a growth rate g with a factor c, where $c > 1$, corresponds to a larger objective function value than a correction with $1/c$ especially if c is large. The objective function values are $\left((c-1)g\right)^2$ and $\left(\frac{(c-1)}{c}g\right)^2$ respectively. Since large upward corrections from a close-to-zero value are relatively costly, these are avoided as much as possible. Thus, the GRP's benchmarked values move more gradually from a close-to-zero value than Denton's results do. To compensate for this, larger peaks may be necessary for the following time-periods to fulfill the temporal aggregation constraint. As benchmarking usually aims at as smooth as possible corrections over time, irregular peaks can be considered undesirable. Related to the relatively slow growth from a close to zero value is that the peaks tend to turn up later in time than for a time-symmetric method like Denton PFD. For the backward variant of GRP the opposite occurs, benchmarked time series move relatively quickly from a close to zero value, which gives rise to relatively early peaks. The example in Subsection 3.3.3 illustrates this problem.

3.3.3 Example

Below we present an example that illustrates the problems of GRP methods. In this example, a time series consisting of 15 months is reconciled with 5 quarterly values. The monthly series is constant: each monthly value is 10. The quarterly values are: 80, 250, 80, 400 and 100, respectively. Figure 3.1 compares the results of Denton PFD, GRPF and GRPB.

As the largest differences occur between both GRP methods, time reversibility is obviously not satisfied. The highest and lowest points appear at different months. The example clearly shows that the use of a different benchmarking method may lead to substantially different conclusions.

In accordance with Subsection 3.3.2, GRPF leads to relatively late peaks, i.e. at the last month of each quarter, while GRPB results in early peaks, i.e. at the first month of each quarter. Denton PFD's results are in between, peaks and troughs occur at the middle month of each quarter.

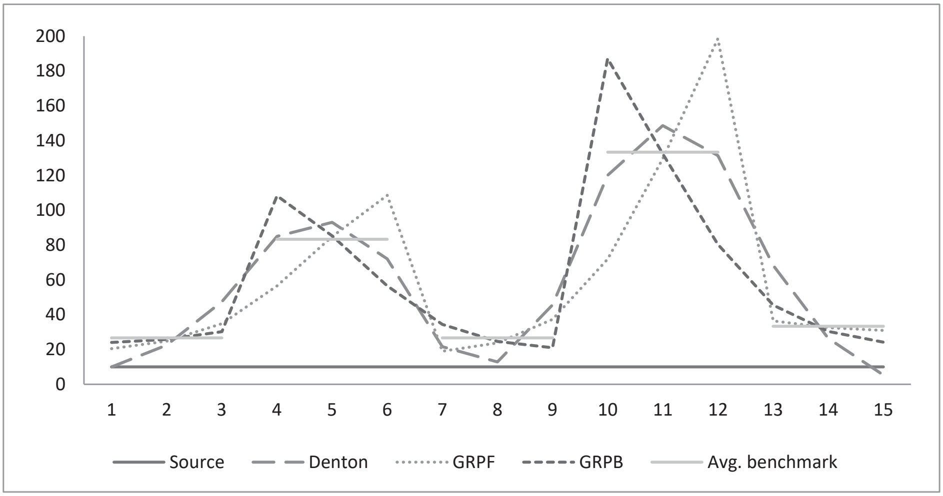

Figure 3.1 example: results of three benchmarking methods. 'Avg. benchmark' stands for the average level of the monthly values that complies with the quarterly benchmarks and that is computed as one-third of its quarterly counterpart.

It needs however to be noted that the example cannot be considered representative for real life applications. In general, benchmarking methods are not meant to be used for reconciling large differences and for constant sub annual series. To explain the latter, a main assumption of Denton PFD is that the sub annual series provides information about short-term change. Constant series however cannot be considered very informative. Nevertheless, the problem of reconciling constant term series does occur in problems that are closely related to benchmarking, like interpolation and calenderization (see e.g. Dagum and Cholette, 2006 and Boot *et al.* 1967).

The reason for choosing this example is purely educational. It provides good insight into properties of the different types of objective functions. The reader is referred to Subsection 3.5.3 for more realistic examples.

3.4 ALTERNATIVE BENCHMARKING TECHNIQUES

In Section 3.3 we identified two problems with GRP methods. In this section we consider two alternative benchmarking techniques that solve the time irreversibility property.

3.4.1 Simultaneous growth rate preservation

Here, we propose two alternative objective functions for GRP. The first is a 'time symmetric' variant of GRP, defined by

$$f_S^{GRP}(\boldsymbol{x}) = \frac{1}{2}\sum_{t=2}^{n}\left(\frac{x_t}{x_{t-1}} - \frac{p_t}{p_{t-1}}\right)^2 + \frac{1}{2}\sum_{t=2}^{n}\left(\frac{x_{t-1}}{x_t} - \frac{p_{t-1}}{p_t}\right)^2, \tag{3.5}$$

where subscript S stands for 'simultaneous'. The method will be called GRPS in the remainder of this chapter. The GRPS objective function both preserves forward and backward growth rates. As far as the authors know this method has not been mentioned elsewhere in the literature. It can be easily seen that GRPS satisfies time reversibility: interchanging t and t-1 does not alter the objective function.

However, the second problem in Section 3.3 (singularity of objective function) is not considered. One of the consequences, negative benchmarked values, can be avoided by imposing lower bounds of zero on the benchmarked values. This can be done by including inequality constraints to an optimization problem, which is a well-established technique (e.g. Nocedal and Wright, 2006). The other problems related with singularity can however still occur.

3.4.2 Logarithmic growth rate preservation

Another 'time symmetric' variant of GRP is given by the logarithmic form:

$$f_L^{GRP}(\boldsymbol{x}) = \sum_{t=2}^{n}\left[\log\left(\frac{x_t}{x_{t-1}}\right) - \log\left(\frac{p_t}{p_{t-1}}\right)\right]^2. \tag{3.6}$$

This function was firstly considered by Helfand *et. al.* (1977). It is immediately verified that function (3.6) satisfies the time reversal property as well. The objective function can be considered the logarithmic version of GRP and equally well the logarithmic version of Denton PFD. It will be denoted GRPL in the remainder of this chapter, where *L* stands for 'logarithmic'.

Note that (3.6) can be used for strictly positive preliminary values only, and that it produces benchmarked values that are larger than zero as well. This does not seem an important limitation, as Section 3.3 already mentioned that growth rate preservation can be considered inappropriate for problems with positive and negative values. Nevertheless, a potential solution for time series with negative values is to add a sufficiently large constant to the series prior to benchmarking and subtract that constant from the benchmarked series. A potential drawback of this solution is that adding a constant distorts initial growth-rates. Thus, it is unclear whether preliminary growth rates are actually preserved. Further research is necessary to better understand the implications of this solution.

Although GRPL necessarily produces positive values, other problems in Section 3.3.2, related to a singular objective function can still occur.

3.4.3 Comparison

When comparing GRPS and GRPL, it can be expected that GRPL behaves more like Denton PFD. Below we give two reasons for this.

Firstly, because of the asymptotic properties of the log function, the problem that close-to-zero values are avoided is less severe for GRPL than for GRPS. Close-to-zero values are associated with large adjustments of growth rates. Very large adjustments of growth rates are penalised less in GRPL than in GRPS, since GRPS's objective function grows faster when corrections are large.

Secondly, a first-order Taylor linearization of GRPL's objective function corresponds to Denton PFD's function, whereas the approximation of GRPS leads to a different result. The first-order Taylor linearization of a bivariate function $f(x_1, x_2)$ in (x_1^0, x_2^0) is given by $f(x_1^0, x_2^0) + f'_{x1}(x_1^0, x_2^0)(x_1 - x_1^0) + f'_{x2}(x_1^0, x_2^0)(x_2 - x_2^0)$. It follows that the linearization of the squared terms of the GRPL's and GPRS's objective functions

$$\log\left(\frac{x_t}{x_{t-1}}\right) - \log\left(\frac{p_t}{p_{t-1}}\right)$$

and

$$\frac{1}{2}\left(\frac{x_t}{x_{t-1}} - \frac{p_t}{p_{t-1}}\right) + \frac{1}{2}\left(\frac{x_{t-1}}{x_t} - \frac{p_t}{p_{t-1}}\right)$$

in the preliminary values

$$(p_t,\ p_{t-1})$$

are given by

$$\left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)$$

and

$$\frac{1}{2}\left(\frac{p_t}{P_{t-1}} + \frac{p_{t-1}}{p_t}\right)\left\{\left(\frac{x_t}{p_t}\right) - \left(\frac{x_{t-1}}{p_{t-1}}\right)\right\},$$

respectively.

Example

In order to explore the properties of GRPL and GRPS, we consider the example of Subsection 3.3.3 again. Figure 3.2 compares results of the symmetric f_S^{GRP}, f_L^{GRP} and $f^{PFD}(\boldsymbol{x})$ methods.

Firstly, it can be seen that the peaks and troughs occur at the same periods for all time symmetric methods. Secondly, some of the drawbacks related to the singularity of the objective function still occur. When compared to Denton PDF, GRP methods tend to avoid close-to zero values, move away relatively slowly from low values (in both directions) and lead to irregular large peaks. Thirdly, in accordance to Subsection 3.4.3, GRPL resembles Denton PFD more than GRPS, which follows from the slightly lower peaks of GRPL.

3.5 EMPIRICAL TEST

In this section an illustration exercise is conducted on real-life data, in order to find out whether or not the problems mentioned in Section 3.3 do occur in a realistic, practical application.

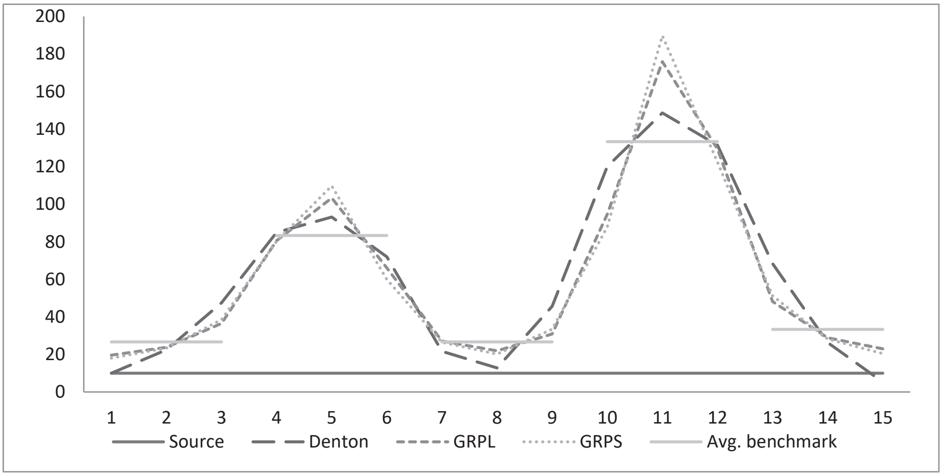

Figure 3.2 example: results of three symmetric benchmarking methods. 'Avg. benchmark' stands for the average level of the monthly values that complies with the quarterly benchmarks and that is computed as one-third of its quarterly counterpart.

3.5.1 Data Sets

The data set used for the illustration is obtained from quarterly and annual trade as published on the website of United Nations (UN).

The United Nations Commodity Trade Statistics Database (UN Comtrade) contains data from statistical authorities of reporting countries, or are received via partner organizations like the Organisation for Economic Co-operation and Development (OECD). The United Nations Totaltrade (UN Tottrade) data are mostly taken from the International Financial Statistics (IFS), published monthly by the International Monetary Fund (IMF). Differences between both sources emerge because of differences in data collection methods and purposes (United Nations, 2017). All data are publicly available at http://comtrade.un.org/.

We use UN Tottrade as data source for quarterly data and both UN Tottrade and UN Comtrade as sources for annual data. Both data sources include imports and exports for approximately 200 UN member states.

For our application all series were selected that include three annual totals and twelve quarterly values for 2002-2004. The variables of interest are total imports and exports. Series with quarterly or annual values smaller than 0.1 million dollars were deleted, as multiplicative benchmarking methods cannot be considered appropriate for zero or near zero values (see Subsection 3.2). Since the series are in million dollars, the cutoff value only excludes "extreme" cases and still leaves some real-life cases of singularity issues.

We end up with 238 time series for Comtrade and 253 series for Tottrade. The average year to year growth rates discrepancy between the annualized quarterly series and their benchmarks are 5.9%-point and 2.7%-point for Comtrade and Tottrade benchmarks, respectively. For the majority of series the discrepancy can be considered small. The percentage of series with a maximum discrepancy below 5%-point are 79% and 87%, respectively.

3.5.2 Results

Our first aim is to assess overall performance. We will compare the degree of preservation of the preliminary values and their growth rates for the various methods that are discussed in this chapter.

Table 3.1 shows for the five methods the median values over all series, for the functions $f_F^{GRP}, f_B^{GRP}, f_S^{GRP}$ for forward, backward and simultaneous movement preservation and f^{Level} for preliminary value preservation. The latter function measures total squared relative adjustment, defined by

$$f^{Level}\left(\boldsymbol{x}\right) := \sum_{t=1}^{n} \left(\frac{x_t}{p_t} - 1\right)^2. \tag{3.7}$$

The rows of Table 3.1 display the methods and the columns show the median value for four evaluation criteria. For instance, the result in the second row and the third column is a median value for the simultaneous growth rate preservation criterion f_S^{GRP}, as defined in (3.5).

It can be seen from Table 3.1 that the GRPF method, that is designed to preserve forward growth rates, results in relatively poor backward movement preservation. The opposite is also true: GRPB does not preserve forward movements very well. From these results, we can conclude that time reversibility actually matters. Table 3.1 also demonstrates that the time symmetric methods, Denton PFD, GRPS and GRPL, perform well on all measures and that difference between those methods are only marginal.

To assess forward, backward and simultaneous growth rate preservation, a relative criterion is used that compares the values of the objective functions $f_F^{GRP}\left(\boldsymbol{x}\right)$, $f_B^{GRP}\left(\boldsymbol{x}\right)$ and $f_S^{GRP}\left(\boldsymbol{x}\right)$ with their optimum values, which are obtained from GRPF, GRPB and GRPS, respectively.

Table 3.1 Median values of criteria in (3.1), (3.3), (3.5) and (3.7)

	COM data set				TOT data set			
	f_F^{GRP}	f_B^{GRP}	f_S^{GRP}	f^{Level}	f_F^{GRP}	f_B^{GRP}	f_S^{GRP}	f^{Level}
Denton PFD	0.87	0.88	0.88	26.42	0.33	0.41	0.37	2.07
GRPF	0.84	0.98	0.93	26.43	0.27	0.48	0.45	2.06
GRPB	1.00	0.82	0.91	26.47	0.48	0.28	0.45	2.07
GRPS	0.87	0.89	0.88	26.41	0.34	0.38	0.36	2.07
GRPL	0.87	0.88	0.88	26.42	0.33	0.41	0.37	2.07

The values for the COM and TOT data sets are $*10^{-2}$ and $*10^{-5}$, respectively.

Analogous to the standards in Di Fonzo and Marini (2012a), movement preservation is considered acceptable if it lies within 10% of the optimum value. That is, if $f^{\,method}\left(\boldsymbol{x}\right)/f^{\,optimum}\left(\boldsymbol{x}\right) \leq 1.1$, where f is one of the previously mentioned objective functions.

For the five methods considered, Table 3.2 shows the percentage of time series with acceptable forward, backward and simultaneous movement preservation.

Table 3.2 Percentage of time series with acceptable movement preservation

	COM data set			TOT data set		
	Forward	**Backward**	**Simult.**	**Forward**	**Backward**	**Simult.**
Denton PFD	79.4	78.6	95.8	79.4	79.4	96.0
GRPF	100.0	48.7	81.5	100.0	47.8	82.6
GRPB	47.1	100.0	76.9	44.3	100.0	75.1
GRPS	82.4	77.3	100.0	80.6	79.4	100.0
GRPL	79.8	79.0	96.6	79.4	79.4	96.0

For Denton PFD an acceptable degree of simultaneous movement preservation is found for more than 95% of all cases. Thus, one can conclude that Denton PFD can be considered as a very good approximation for the optimal GRPS method; the approximation is even better than the GRPF and GRPB methods, for which acceptable performance is found for around 80% of all cases.

So far, we focused on performance for entire time series. Below we will consider the occurrence of large and extreme reconciliation adjustments made to single values and growth rates.

To measure the adjustments made to growth rates, the absolute difference $|g_{it}(\boldsymbol{x}) - g_{it}(\boldsymbol{p})|$*100% is used, where g_{it} is a growth rate for series i and period t. Tables 3.3 and 3.4 compare the occurrence of large and extremely large adjustments to forward, backward and simultaneous growth rates.

Table 3.3 Percentage of large growth rate adjustments (> 10%-point difference)

	COM data set			TOT data set		
	Forward	**Backward**	**Simult.**	**Forward**	**Backward**	**Simult.**
Denton PFD	2.0	2.1	1.9	0.8	0.6	0.6
GRPF	1.9	2.4	2.3	0.6	0.9	0.7
GRPB	2.3	1.5	2.0	1.1	0.3	0.8
GRPS	1.9	1.9	1.8	0.8	0.6	0.6
GRPL	2.0	1.9	1.9	0.8	0.6	0.5

These tables show minor differences. Small differences between methods are also in observed in Table 3.5, which shows large and extreme corrections to preliminary values, as measured by the relative criterion (x_{it}/p_{it})*100%. Hence, one can conclude that the problems caused by singularity do not translate into more often occurring large corrections for the data set under consideration.

Table 3.4 Percentage of extreme growth rate adjustments (> 50%-point difference)

	COM data set			TOT data set		
	Forward	**Backward**	**Simult.**	**Forward**	**Backward**	**Simult.**
Denton PFD	0.3	0.2	0.4	0.1	0.1	0.1
GRPF	0.2	0.2	0.2	0.0	0.1	0.1
GRPB	0.3	0.0	0.2	0.2	0.0	0.1
GRPS	0.2	0.1	0.2	0.1	0.0	0.1
GRPL	0.2	0.1	0.2	0.1	0.0	0.1

Most remarkable in Table 3.5 are the negative benchmarked values obtained for GRPF in the TOT data. An example of this is illustrated in Figure 3.5.

Table 3.5 Percentage of large adjustments to preliminary values

	COM data set			TOT data set		
	Large (>10%)	**Extreme (>100%)**	**Negative (<0%)**	**Large (>10%)**	**Extreme (>100%)**	**Negative (<0%)**
Denton PFD	13.2	1.0	0.0	5.8	0.4	0.0
GRPF	13.0	1.0	0.0	5.8	0.3	0.1
GRPB	13.1	0.9	0.0	5.6	0.3	0.0
GRPS	13.1	0.9	0.0	5.8	0.4	0.0
GRPL	13.0	0.9	0.0	5.8	0.4	0.0

Despite the similar results of the five benchmarking methods in Tables 3.3-3.5, there are clear differences in smoothness of reconciliation adjustments. To demonstrate this, we will use the smoothness indicator (Temurshoev, 2012).

$$\text{Smoothness} = \sum_{t=2}^{n-2}\left[BI_t - \bar{BI}_t\right]^2,. \qquad (3.8)$$

where BI_t is the so-called benchmark-to-indicator ratio, i.e. x_t/p_t and $\overline{BI}_t$ is the 5-terms moving average $\frac{1}{5}\sum_{k=t-2}^{k=t+2} BI_k$.

According to this indicator, we find in Table 3.6 that the smoothest results are obtained for Denton PFD and GRPL. Conversely, the asymmetric GRPF and GRPB methods yield the most irregular adjustments. It follows that the time-symmetric method GRPS, but most so GRPL, suffers less from singularity than the asymmetric methods GRPF and

GRPB do. These results most clearly illustrate the problems with the singularity of GRP's objective function that were described in Subsection 3.3.2.

Table 3.6 Smoothness indicator values (3.8), summed over all series

	COM data set	TOT data set
Denton PFD	3.4	0.3
GRPF	9.8	39.0
GRPB	8.2	2.9
GRPS	4.3	1.1
GRPL	3.3	0.5

3.5.3 Examples

Below we show two examples to demonstrate that the problems in Section 3.3 do occur in a real-life application.

The first example, in Figures 3.3 and 3.4, illustrates that non-symmetric GRP methods may change the timing of the most important economic events. When considering the first nine quarters, the two highest values occur at different time periods. GRPF's peak periods are at quarters 6 and 7 and those of GRPB are at quarters 5 and 6. Closely related to this, is that GRPF moves away relatively slowly from the close-to-zero values at quarters 1-4.

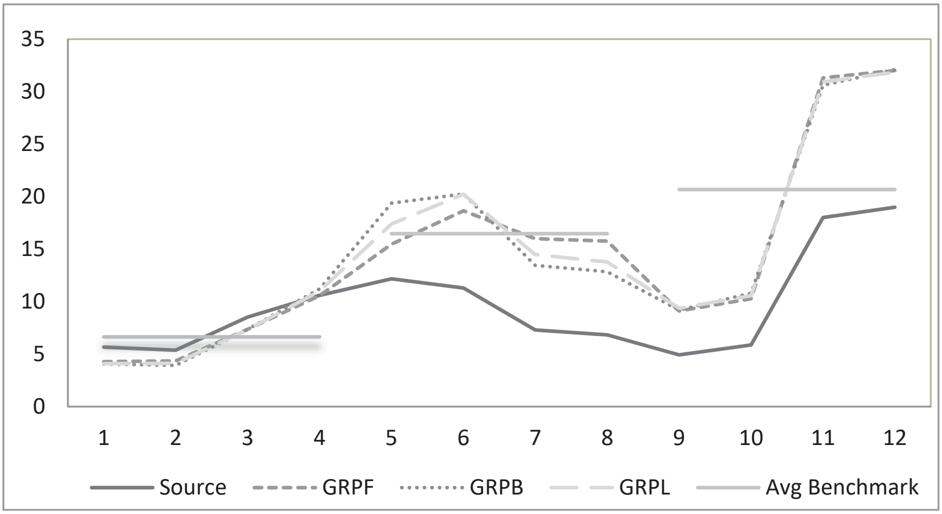

Figure 3.3 Exports Burundi, Comdata, 2002-2004, millions of US dollar. 'Avg Benchmark' stands for the average level of the quarterly data that complies with the annual benchmarks and that is computed as one-fourth of its annual counterpart.

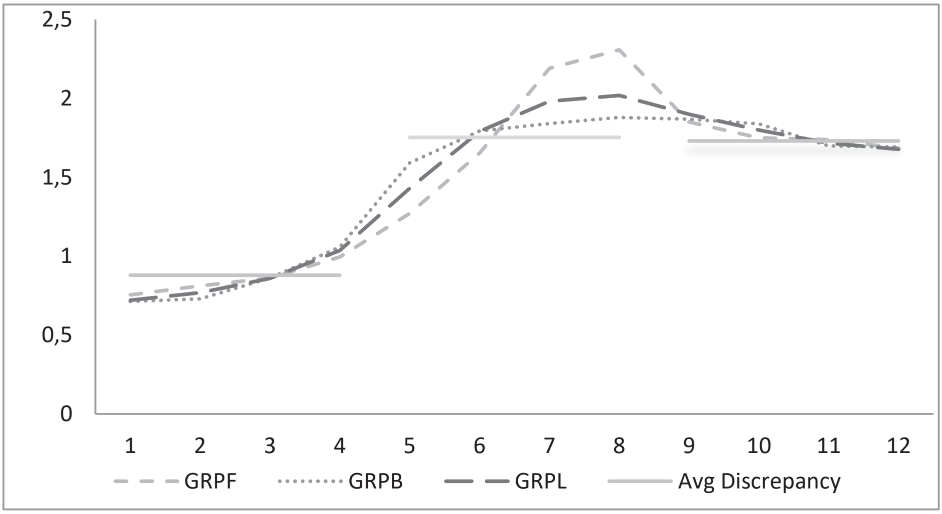

Figure 3.4 Benchmark to Indicator ratios, Exports Burundi, 2002-2004. 'Avg Discrepancy' stands for the annual BI-ratio, i.e. the ratio of an annual benchmark and the sum of the underlying quarterly indicators.

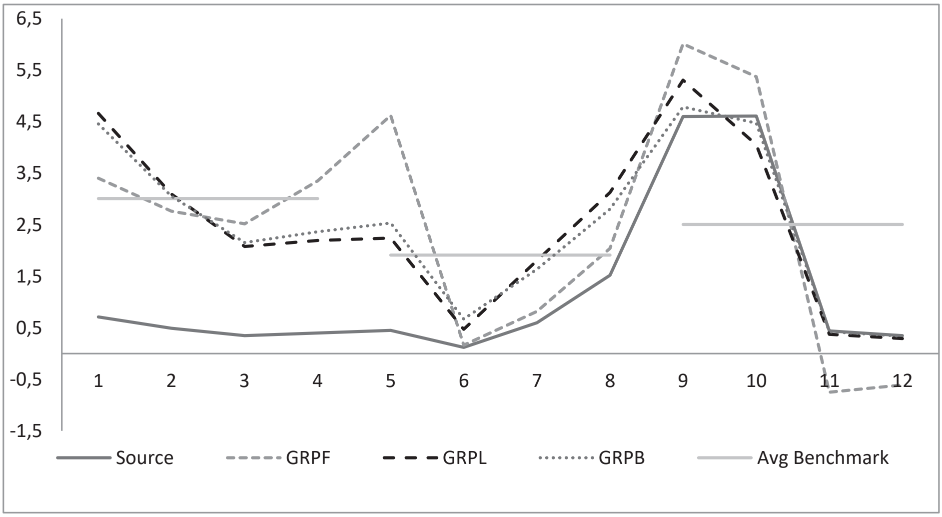

Figure 3.5 Exports Gambia, Totdata, 2002-2004, millions of US dollar 'Avg Benchmark' stands for the average level of the quarterly data that complies with the annual benchmarks and that is computed as one-fourth of its annual counterpart.

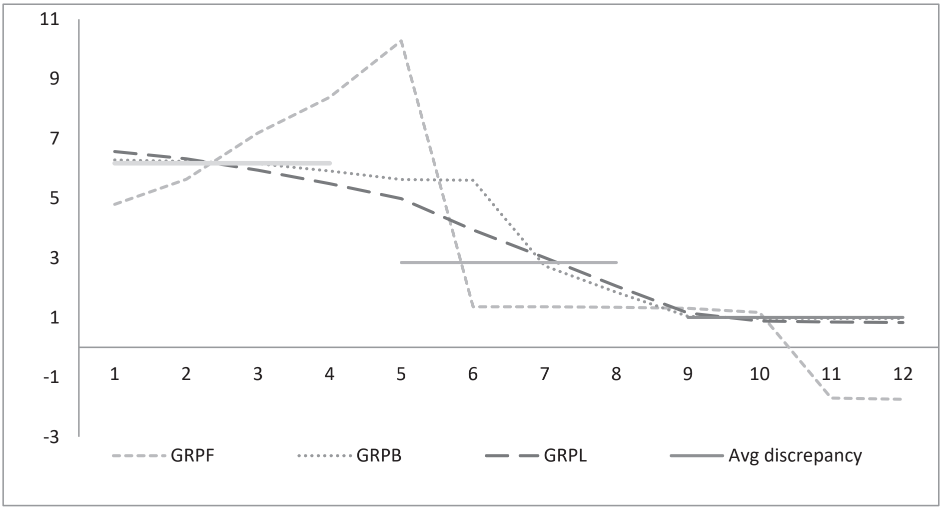

Figure 3.6 Exports Gambia, Totdata, 2002-2004, benchmark to indicator ratio 'Avg Discrepancy' stands for the annual BI-ratio, i.e. the ratio of an annual benchmark and the sum of the underlying quarterly indicators.

The second example illustrates the complications of a singular objective function. As shown in Figure 3.6, GRPF closely preserves growth rates of the quarters 6-10. This comes however at the expense of an irregular peak in quarter 5 and negative benchmarked values in the quarters 11 and 12.

3.6 CONCLUSIONS

Two well-known multiplicative benchmarking methods are Denton Proportionate First Differences (PFD) and Growth Rates Preservation (GRP). It is generally agreed that GRP has the strongest theoretical foundation. It better preserves initial growth rates than Denton PFD. However, from a technical point of view, Denton is the easiest method to apply. Because of this, and because Denton PFD is often a good approximation of GRP, Denton PFD is more popularly applied.

In this chapter two drawbacks of GRP are demonstrated that, to the best knowledge of the authors, have not been mentioned elsewhere.

The first drawback is that GRP does not satisfy the time reversibility property. According to this property it should not matter for the results whether forward or backward growth rates are preserved. That is, benchmarking an original time series, $t = 1, \ldots, n$, or a 'reversed' time series, $t = n, \ldots, 1$ should lead to the same benchmarked series. Since direction of time is irrelevant for any benchmarking application, any benchmarking method should preferably satisfy time reversibility. Moreover, a benchmarking method that does not satisfy time reversibility may yield entirely difficult conclusions on the timing of economic events depending on the chosen time direction. For these reasons forward and backward GRP methods should preferably be discouraged.

In this chapter two alternative GRP methods are presented that do satisfy time reversibility. The first alternative, a new GRPS method, preserves both forward and backward growth rates. The other alternative, an existing GRPL method, preserves logarithms of the forward growth rates.

A second drawback of all GRP methods in this chapter are the singularities in its objective functions. Complications of this are: avoidance of close to zero outcomes, irregular peaks in results and unnecessary negative values in benchmarked results.

These problems actually occurred in an illustrative application on real-life data. Although unnecessary negative values only occasionally occurred, reconciliation adjustments are much more irregular than for Denton PFD. Since smoothness of reconciliation adjustments (BI ratios) is often the main interest of benchmarking, asymmetric GRP methods can be discouraged for many applications.

While the literature considers Denton PFD 'a good approximation' of the ideal GRP method, our main conclusion is that Denton PFD is even more appropriate than standard GRP for many applications. Denton is computationally easier to apply, it does not suffer from the problems related to time irreversibility and a singular objective function. Furthermore, the approximation of Denton PFD's results is even more close for the time-symmetric versions of GRP than for standard GRP.

However, when growth rate preservation is the key point of interest, a time-symmetric version of GRP can also be a good choice, most in particular GRPL. Time symmetric methods preserve growth rates slightly better than Denton PFD, satisfy time reversibility and suffer less severe from the drawbacks of a singular objective function than standard GRP.

4. On the sequential benchmarking of sub-annual series to annual totals[3]

Summary. Temporal benchmarking according to Denton (1971) is widely used for official statistics. The purpose of Denton methods is to achieve consistency between high and low frequency data, e.g. quarterly and annual data. The high frequency data are adjusted to align with low frequency data, while preserving as much as possible the short-term movements of the preliminary high frequency data. Theoretically, it is best to benchmark all available data for all periods at once. Practically, such a simultaneous approach is often not feasible, due to the impossibility of changing results that have already been published. Therefore, benchmarking is often applied according to a sequential approach. This paper demonstrates that a popular Denton method is not always appropriate for sequential benchmarking. Undesirable, abrupt changes of benchmarking corrections can occur. This paper proposes solutions that better preserve the short-term movements between sequentially benchmarked series.

3 This chapter has been published as: Daalmans J. (2018), On the sequential benchmarking of subannual series to annual totals. *Statistica Neerlandica*, 72: 406–420

4.1 INTRODUCTION

Benchmarking is the problem of achieving numerical consistency between time series data measured at different frequencies, e.g. annual and sub annual series. Benchmarking monthly and quarterly series to annual totals is common practice for many National Statistical Institutes (NSI's). For example, each year Statistics Netherlands aligns 12 quarterly Supply and Use Tables with the three most recent annual accounts (Eurostat, 2013, Annex 8C).

Usually, the low frequency (LF) data describe levels and long-term trends better than the high frequency (HF) sources. The latter, on the other side, provide the only information on short-term movements. Most benchmarking methods combine the strengths of both sources. When achieving consistency, the LF benchmarks are fixed, while preserving as much as possible the one- period ahead movements of the HF series. The latter property is also known as 'movement preservation principle'. Movement preservation implies that reconciliation adjustments are dependent in time. The adjustment for the current period depends on the adjustments for the previous and the following periods. In this way the so-called step-problem is avoided, i.e. spuriously large distortion of growth rates between two consecutive HF periods in different LF periods, e.g. January to December.

Several benchmarking methods are available in the literature. The Proportional First Differences (PFD) Denton method (Denton, 1971) is particularly popular. Benchmarking methods with movement preservation, like Denton, have the property that an incorporation of new data at the end of a time series may yield different benchmarked results for all previous HF periods. Although the impact of this gradually diminishes for distant periods in the past (Bloem *et al.*, 2001), this property can be cumbersome in practice, because there is limited possibility of adjusting results that have already been published. Usually, a revision policy is applied that only allows the data for the most recent years to change. For this reason, benchmarking is often applied to relatively short time-series. For example, the NSI's of the USA (Chen and Andrews, 2008, p33) and the Netherlands (Eurostat, 2013, Annex 8C) apply benchmarking to time series of three years.

A drawback of this sequential approach is that more reconciliation adjustment may be made than when all periods are reconciled at once. In particular, there is a risk of large discontinuities of growth rates between two consecutive HF periods in two different benchmarked series; a problem that has similarities with the previously mentioned step problem.

Although a solution to this problem has already been proposed in Denton (1971), the sequential benchmarking problem has not been mentioned often in the literature. The focus of the current chapter is on a particular sequential benchmarking problem, in which the reconciliation adjustments do not monotonically increase or decrease around the end of the estimation intervals. This chapter's aim is to demonstrate that the popularly applied Denton methods are not always appropriate for this particular problem. Alternative methods will be proposed and it will be demonstrated that these methods improve on an existing Denton method.

Benchmarking problems with non-monotonic adjustment patterns are regularly mentioned in the literature. The existing literature seems to suggest that the problem is more relevant for month-to-quarter benchmarking than for other combinations of LF and HF periods, like month-to-year or quarter-to-year. For example, Delden and Scholtus (2017) present an application to short-term business statistics in which monthly turnover data, as directly collected by Statistics Netherlands, are benchmarked to quarterly totals, that are largely based on Value Added Tax (VAT) registers. For the majority of industries, the benchmark values for all first quarters of a year are higher than the sum of the underlying monthly values, while the opposite was found for the fourth quarters of the year. Part of these differences was explained from reporting patterns in the VAT data. Another example of month-to-quarter benchmarking is described by Battista *et al.* (2009) for benchmarking USA's employment statistics. Structural correction patterns were explained by notable, but not understood, differences in seasonal patterns of the monthly and quarterly data. Di Fonzo and Marini (2012b) provide several examples of quarter-to-year benchmarking. In most of their applications benchmarking corrections show regular adjustment patterns. Usually, there is a monotonic increase or decrease within a certain interval of time. This suggests that the newly proposed methods in the current chapter are not very useful for the examples in the Di Fonzo and Marini paper.

The structure of this chapter is as follows. Section 4.2 presents a sequential approach to benchmarking that can be applied with the currently available Denton method. A drawback of this approach is explained in Section 4.3. Section 4.4 proposes alternative methods that are suitable for specific problem classes. Empirical applications are given in Section 4.5. Section 4.6 presents generic solutions that can be applied to any data set.

4.2 CURRENT APPROACH FOR SEQUENTIAL BENCHMARKING

Subsection 4.2.1 gives a formal description of a Denton method. Subsection 4.2.2 explains how the Denton method can be properly applied to a sequential benchmarking problem.

4.2.1 General notation and Denton PFD

Let p_t, $t = 1,\ldots,n$ be a preliminary HF series. The LF benchmark series are denoted by b_T, $T = 1,\ldots,N$. Assume that $n = sN$, where s is an integer, which reflects the aggregation order. For instance, s equals to 4 for quarter-to-annual aggregation and $s = 3$ for month-to-quarter aggregation. The benchmarking problem is to find a vector of benchmarked values x_t, $t = 1,..,n$ that satisfies the temporal aggregation constraints.

Several types of aggregations constraints are described in literature, see e.g. Dagum and Cholette (2006, Subsection 3.5). For ease, we only consider the type of constraint in which a sum of HF values needs to match one LF value; a constraint that can be expressed as $\sum_{t \in T} x_t = b_T$, $T = 1, \ldots, N$, where $t \in T$ refers to the HF periods that belong to LF period T. Usually, we have $\sum_{t \in T} p_t \neq b_T$, otherwise no correction would be needed.

In matrix form a set of temporal aggregation constraints can be expressed as a linear system of equalities $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, in which $\mathbf{A}$ is a $N \times n$ matrix converting high- into low-frequency values. This matrix is defined by

$$\mathbf{A} = \mathbf{I}_N \otimes \boldsymbol{a}^T, \tag{4.1}$$

where $\otimes$ is the Kronecker product and $\boldsymbol{a}$ is the $(s \times 1)$-vector given by $\boldsymbol{a} = 1_s = (1, \ldots, 1)^T$. For example, the matrix $\mathbf{A}$ for reconciling eight quarters with two annual totals is given by

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{4.2}$$

The vector $\boldsymbol{b}$ contains the low frequency benchmarks.

Denton (1971) proposed a benchmarking procedure grounded on the Proportionate First Differences (PFD) between the target and the original series. The PFD benchmarked estimates are obtained as the solution to the constrained quadratic minimization problem

$$\min_{\boldsymbol{x}} f^{PFD}(\boldsymbol{x}) \text{ subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \tag{4.3}$$

where

$$f^{PFD}(\boldsymbol{x}) = \left(\frac{x_1}{p_1} - 1\right)^2 + \sum_{t=2}^{n} \left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)^2. \tag{4.4}$$

The second term of the objective function $\sum_{t=2}^{n} \left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)^2$ reflects the movement preservation principle. It attempts to keep the relative adjustments as constant as possible over time, which has been shown to be approximately the same as preserving one-period ahead growth rates, e.g. Cholette (1984). The first term of $f^{PFD}(\boldsymbol{x})$ is based on the assumption that before the current estimation interval a previous period exists that has already been benchmarked. It is derived from $\left(\frac{x_1}{p_1} - \frac{x_0}{p_0}\right)^2$,where x_0 and p_0 refer to the last HF period of the preceding estimation interval. This term aims to preserve the change of the preliminary series between the periods 0 and 1. Since period 0 has already been benchmarked, it follows that $x_0 = p_0$, yielding $\left(\frac{x_1}{p_1} - 1\right)^2$in (4.4).

Cholette (1984) slightly modified Denton PFD by proposing an objective function $f^{CHO}(\mathbf{x})$ that ignores the link with period 0.

$$f^{CHO}(\boldsymbol{x}) = \sum_{t=2}^{n} \left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)^2. \tag{4.5}$$

Some authors consider the Cholette modification as an improved version of Denton (e.g. Di Fonzo and Marini, 2013). This should however only apply if there is not a previous period, or if the link with the previous period is not relevant. As the sequential benchmarking of time series is our main interest, this chapter focuses on Denton PFD rather than on the Cholette modification.

4.2.2 Sequential benchmarking with Denton

The purpose of this subsection is to describe a strategy for the sequential benchmarking of time series. To obtain a benchmarked series that forms one whole for all periods, it will be explained that the set of preliminary values should be rescaled at the start of each benchmarking application. The rescaled series $\tilde{p}_t$ is constructed as follows:

$$\tilde{p}_t = p_t(x_0/p_0) \qquad t = 1, \ldots, n, \tag{4.6}$$

The preliminary values of current benchmarking period are multiplied by the relative correction for period 0, the last period of previous series. By doing this, it follows that $\tilde{p}_1 = x_0 \frac{p_1}{p_0}$, the preliminary estimate for period 1 is connected to the benchmarked value x_0 by multiplication with the observed growth rate $\frac{p_1}{p_0}$. Consequently, preserving the first preliminary value in $f^{PFD}(\boldsymbol{x})$ actually comes down to preserving the growth rate $\frac{p_1}{p_0}$. Not rescaling a preliminary series would imply that the current series is connected to the preliminary value p_0, rather than to x_0. This might introduce a step between the two benchmarked series if x_0 differs from p_0.

This observation was also made by Cholette (1984, Subsection 6.1), who proposed an other solution, based on the inclusion of the difference of p_0 and x_0, in the current period's objective function. Mathematically, Cholette's solution is equivalent to the above-mentioned rescaling procedure. A practical disadvantage of Cholette's method is the need to keep track of the benchmarking corrections of previous estimation interval. After rescaling the preliminary data, the benchmarking problem for the current estimation interval becomes independent of previous corrections.

The rescaling procedure affects the original Denton method, but does not have any influence on the Denton-Cholette modification, as the latter method does not preserve levels. Unless specified otherwise, it will be assumed henceforth that preliminary values are rescaled at the beginning of each sequential benchmarking application. For ease of notation, the tilde in $\tilde{p}_t$ will be omitted.

4.3 DRAWBACK OF CURRENT APPROACH

A drawback of benchmarking according to the sequential process is that there might be unnecessarily large adjustment to the short term movement of the HF series. When benchmarking the current time series, it is not taken into account that there is a sequential time series, whose results depend on the outcomes for the current period. An unfortunate result for the last HF period of current estimation interval might imply unnecessarily large adjustment to the sequential series. The following example illustrates this.

Example

Consider a benchmarking problem consisting of four years and 16 quarters. The time series are benchmarked according to a sequential process, in which two years are processed at a time. The results of the first benchmarking period are depicted in Table 4.1. The benchmarked results are obtained with the Denton method in (4.4).

4

Table 4.1 Results for the first benchmarking period (rounded)

	Q1	Q2	Q3	Q4	Yr. 1	Q5	Q6	Q7	Q8	Yr. 2
HF source	80	60	40	70	250	70	50	40	90	250
LF source	-	-	-	-	225	-	-	-	-	275
Benchmarked	72.6	52.4	35.4	64.7	225	71.1	54.1	45.2	104.6	275

Table 4.2 shows the input for the second estimation interval. As shown in (4.6), the rescaled series is obtained by multiplying the original series by 104.6/90, the ratio of the benchmarked and preliminary value for Quarter 8. The sums for the Quarters 9-12 for the original series and the rescaled series amount to 250 and 290.6, respectively. These sums have to be benchmarked to 225, the value obtained from the annual source. Obviously, more correction is needed for the rescaled series than for the original series. One might therefore conclude that the benchmarked value 104.6 for Quarter 8 is not a good result, when taking Year 3 into account. The benchmarked value is higher than the preliminary value, whereas a downward correction of -10% is required for the following year. The rescaling procedure in (4.6) implies that the best result for Year 3 would be obtained if the benchmarked value for Quarter 8 were 10% lower than the preliminary value, that is 81

Table 4.2 Data for the second benchmarking period (rounded)

	Q9	Q10	Q11	Q12	Yr 3	Q13	Q14	Q15	Q16	Yr 4
HF source	80	60	50	60	250	70	70	40	70	250
HF rescaled	93.0	69.8	58.1	69.8	290.6	81.4	81.4	46.5	81.4	290.6
LF source	-	-	-	-	225	-	-	-	-	250

(90% of 90). As a result, the preliminary values for Quarters 9-12 would be decreased by 10% as well, meaning that no reconciliation adjustment would be necessary for Year 3.

A ratio of a benchmarked and preliminary HF value is often referred to as HF *BI* ratio (benchmark-to-indicator ratio). Similarly, a LF *BI* ratio refers to the ratio of a LF benchmark and the sum of the underlying preliminary HF values. The HF and LF *BI* ratios will be denoted BI_t and BI_T, respectively.

The rescaling procedure in (4.6) implies that the least reconciliation adjustment is obtained for the beginning of a succeeding time series, if $BI_n = BI_{N+1}$, i.e. if the relative benchmarking correction for the current last HF period is the same as the discrepancy for the first LF period of the following series. The problem is however that the required reconciliation adjustment for the future series is not known yet when benchmarking the current series. Therefore, the above-mentioned result cannot be directly used in practice.

As mentioned in Eurostat (2013), with Denton, the benchmarking adjustment for the last HF period, BI_n, largely depends on the adjustments made in the last two LF periods. If there is an upward trend, i.e. if $BI_N > BI_{N-1}$, the relative adjustment for n tends to be larger than the average discrepancy for its corresponding LF period, $BI_n > BI_N$. Thus, Denton anticipates a further increase for $N+1$. In Example 4.1, the relative discrepancy for Year 2 is larger than for Year 1 (275/250= 1.1 versus 225/250= 0.9) and the Denton adjustment for Quarter 8 is indeed larger than the annual discrepancy for the second year (104.6/90= 1.16 versus 1.1).

Using the result that the least reconciliation adjustment is obtained for the succeeding period $N+1$ if $BI_{N+1} = BI_n$, it follows that the least adjustment is obtained, if $BI_N > BI_{N-1}$ goes together with a BI_{N+1} with $BI_{N+1} > BI_N$. A similar conclusion can be drawn for a downward trend; if $BI_N < BI_{N-1}$ Denton implicitly assumes the unknown $BI_{N+1}$to be smaller than BI_N. In brief, Denton assumes that the local trend of adjustments at the end of the current estimation interval continues at the beginning of a sequential series.

The above-mentioned assumption seems to be reasonable for many benchmarking problems, e.g. most of the series in Di Fonzo and Marini (2012b). However, for some applications adjustment patterns are not monotonically increasing or decreasing at the boundaries of the estimation intervals. Such series cannot be expected to be captured well by a sequential application of the Denton method. For example, in the previous example the *BI* ratios for the first three quarters are given by 0.9, 1.1 and 0.9. Since these ratios do not grow monotonically, it follows that Denton PFD can be unsatisfactory for that problem. Alternative methods will be proposed in Sections 4.4 and 4.6.

Closely related to the sequential benchmarking problem is the so called extrapolation problem, which is sometimes known as benchmarking 'forward' series. This problem arises if HF data are available for the most recent periods $n+1$, $n+2$ and $n+3$ and the corresponding LF benchmarks ($N+1$) are missing yet. In addition, there is a need to estimate those LF totals from the underlying HF series and the expected reconciliation adjustment.

The standard assumption in extrapolation is that all future BI ratios are the same as the last observed ratio, BI_n, e.g. Bloem *et al.* (2001). It is however mentioned in Bloem *et al.* (2001) and further elaborated in Di Fonzo and Marini (2012b) that suboptimal predictions are made when the LF *BI* ratios exhibit certain structural patterns, like seasonality.

Di Fonzo and Marini (2012b) propose a model-based solution for the extrapolation problem, based on the available HF data for $n+1$, $n+2$ and $n+3$ and an estimate for the unavailable LF benchmark $N+1$. Their solution cannot be directly applied to sequential benchmarking though, because of the lack of HF data for $n+1$, $n+2$ and $n+3$. Conversely, the new solutions for sequential benchmarking that are proposed in Sections 4.4 and 4.6 can also be applied to the extrapolation problem, but the Di Fonzo and Marini solution has to be preferred for that problem, because their solution makes use of all available HF data up to $n+3$. In conclusion, the proposed solutions in Sections 4.4 and 4.6 can be seen as a modification of the Di Fonzo and Marini method, for the special case that the HF data for $n+1$, $n+2$ and $n+3$ are unavailable.

4

4.4 NEW SOLUTIONS FOR NON-MONOTONIC SERIES

This section presents two new methods for the sequential benchmarking of time series with non-monotonic adjustment patterns and illustrates these solutions by an example.

Method 1

The first new method extends the Denton objective function by including a new term for the level of the last HF value.

$$f^{Meth1}(\boldsymbol{x}) = \left(\frac{x_1}{p_1} - 1\right)^2 + \left(\frac{x_n}{t_n} - 1\right)^2 + \sum_{t=2}^{n}\left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)^2, \tag{4.7}$$

where $t_n = p_n(y_N/\sum_{t\in N} p_t)$ and y_N is the last LF-value.

The additional term $\left(\frac{x_n}{t_n} - 1\right)^2$ measures the squared relative distance between the last HF value and a target value. That target value is the product of the provisional value p_n and the relative discrepancy of its LF period: $y_N/\sum_{t\in N} p_t$. As before, the minimization of (4.7) is subject to the temporal alignment constraints $\boldsymbol{A}\mathbf{x} = \mathbf{b}$.

Method 2

The method is given by

$$\min_{\boldsymbol{x}} f^{PFD}(\boldsymbol{x}) \quad \text{subject to } \widetilde{\mathbf{A}}\boldsymbol{x} = \widetilde{\boldsymbol{b}}, \tag{4.8}$$

where $\tilde{\mathbf{A}}\boldsymbol{x} = \tilde{\boldsymbol{b}}$ includes the temporal aggregation constraints and one additional constraint, given by

$$x_n = p_n(y_N / \textstyle\sum_{t \in N} p_t) \,. \tag{4.9}$$

The added term in the objective function of Method 1, appears as a constraint in Method 2. The two new methods move x_n towards the value that would have been obtained if x_n were subjected to the relative reconciliation adjustment for LF period N. The implicit assumption is that the reconciliation adjustment for $N+1$ is the same as for N. In absence of any knowledge about $N+1$, this can be considered a neutral assumption. The main idea of the new methods is to keep the reconciliation adjustment of x_n in line with the unknown correction for x_{n+1}. Thus, unnecessarily large adjustment for the following estimation interval is meant to be avoided. The more the required adjustment for $N+1$resembles the adjustment for N, the better the results of the methods will be. It should be noted that the extension of the method may come at a cost. The preservation of one additional level implies less room for movement preservation within each estimation interval. So, there can be a trade-off between smooth transitions between successive benchmarked series and movement preservation within individual benchmarked series.

When comparing Methods 1 and 2, Method 1 has the advantage of more flexibility, since x_n is allowed to deviate from its target value. On the contrary, Method 2 has an important practical advantage over various other methods, which is explained below.

Because of the rescaling in (4.6), the preliminary values for the current series depend on the last benchmarked value for previous series. As a result, the benchmarking of the current estimation interval is not possible until the previous interval is done. However, in Method 2 the last benchmarked value of a series is fixed. Thus, the preliminary values for current estimation interval do not depend anymore on the estimation for the previous periods. Hence, all benchmarking problems can be processed independently, thus enhancing the tractability of a sequential process. In Section 4.6 a third method is introduced, which generalizes Method 1.

Example

The two methods have been applied to the example in Section 4.3; in which 8 quarterly values are benchmarked to two annual totals. Figure 4.1 compares the HF BI ratios of the current sequential Denton method (as denoted by "Seq: Denton") with the ratios of the two new methods ("Seq: Method 1" and "Seq: Method 2"). It also displays the results of a simultaneous application of Denton ("Sim: Denton").

At the end of each estimation interval, at quarters 8 and 16, the sequential Denton method continues the upward trend of the previous adjustments. Consequently, the benchmarking adjustments are higher than the annual BI ratio. The reconciliation adjust-

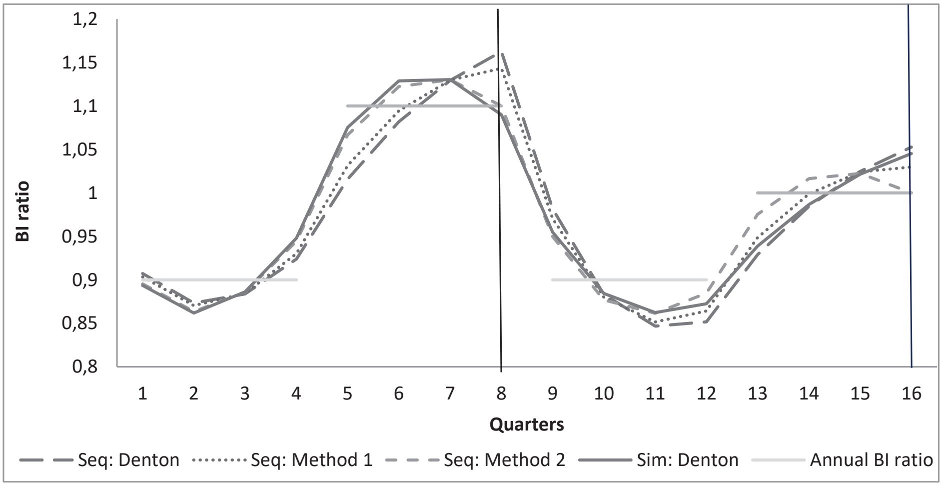

Figure 4.1 Comparison of four benchmarking methods on an illustrative example

ments at the end of the second year strongly contrast with the negative adjustments that are required for the third year, thus implying a need for a sudden change.

For the newly proposed Method 2, the benchmarking adjustments at quarters 8 and 16 are exactly equal to the average annual adjustment. The corrections for the second year are more in line with the required adjustments for the third year. The results for the second and third year are similar to those of simultaneous Denton, the optimal method if all periods can be simultaneously benchmarked. The results for Method 1 are somewhat in between 'sequential Denton' and Method 2.

4.5 EMPIRICAL RESULTS

This section shows the results of an empirical application to three "real-life" problems. Subsection 4.5.1 describes the data sets, results of two evaluation criteria are given in Subsection 4.5.2 and 4.5.3 and Subsection 4.5.4 presents conclusions.

4.5.1 Data

Three data sets are considered. The first two are publically available, the third one cannot be publicly distributed. The main properties of each data set are summarised in Table 4.3.

Data set 1. International trade statistics (Trade)

This data set comes from the World Trade Organisation (WTO, https://www.wto.org). It contains monthly and quarterly data on 'merchandise trade value' and 'trade in commercial services value'. The data include non-seasonal adjusted imports and exports as collected

Table 4.3 Benchmarking problem characteristics

Problem	**Trade**	**Employment**	**Retail**
Number of series	48	53	55
- Non-monotonic	15	23	28
- Monotonic	32	30	27
Number of quarters	12	52	10
Number of months	36	156	30
Number of sequential estimation problems	3	13	5
Average growth rate discrepancy (in %-point)	2.18	0.37	3.55

by approximately 120 member states. From this data set, 220 time series were pre-selected with available data for all 36 months and 12 quarters for 2014-2016. Many time series, with hardly any temporal discrepancy were omitted, as these series are not interesting. The omitted series are the ones for which the minimum and maximum quarterly BI ratio differ less than 0.01. The ultimately selected data include 48 time series.

Data set 2. USA Employment statistics (Employment)

This data set contains monthly and quarterly data on total nonfarm employees from the current employment statistics (CES) and the Quarterly Census of Employment and Wages (QCEW).

The CES is a timely monthly survey that covers approximately one-third of the nation's nonfarm employers. The QCEW contains monthly and quarterly employment data from states' unemployment insurance tax records. It covers almost all nonfarm employers and becomes available later than CES. The aim of the benchmarking exercise is to align the quarterly sums of three monthly CES values with the corresponding benchmarks from QCEW. A data set was constructed containing monthly CES data and quarterly aggregated monthly QCEW data on total employment for 53 US states. The data set covers all 156 months and 52 quarters for the period 2004-2016.

Data set 3. Netherlands Retail statistics (Retail)

The third data set contains monthly and quarterly turnover data for retail services and a range of underlying industries. The monthly data are directly collected by Statistics Netherlands, the quarterly data are largely based on tax registers. A data set is available for 55 industries, covering 30 months and 10 quarters for the period 2015Q1 – 2017Q2.

For trade and employment twelve months are benchmarked to four quarters within each estimation problem. For retail six months and two quarters are processed at each step.

The time series in each data set are divided into two classes: monotonic series and non-monotonic series. The first class includes all series for which in more than half of all cases the quarterly BI ratios monotonically change around the end of the interval, i.e. $BI_{N+1} \geq BI_{N+1} \geq BI_{N-1}$ or $BI_{N+1} \leq BI_{N+1} \leq BI_{N-1}$. All other series are considered non-monotonic.

Table 4.3 shows that about one-third to one-half of all series belong to the "non-monotonic class" , these are series for which the two new methods have been especially developed.

The 'average growth rate discrepancy' in the last row of the table measures the difference of quarterly growth rates between quarterly aggregated monthly data and the quarterly benchmarks. For employment this discrepancy is much smaller for the other two data sets.

4.5.2 Results: Movement preservation

This subsection compares the degree of overall movement preservation of three sequential methods: Denton, Method 1 and Method 2 with the optimal movement preservation from simultaneous Denton. Overall movement preservation is measured by the Denton-Cholette objective function value $f^{CHO}(\boldsymbol{x}_{\text{Meth}})$, as defined in (4.5). The evaluation criterion is defined by $\frac{f^{CHO}(\boldsymbol{x}_{seq.\ meth})}{f^{CHO}(\boldsymbol{x}_{sim.Denton})}$, where $\boldsymbol{x}_{\text{seq. meth}}$ is the result of a sequential method and $\boldsymbol{x}_{\text{sim.Denton}}$ is the outcome of simultaneous Denton. This relative criterion measures how much worse a sequential approach is as compared to simultaneous Denton. Its value is by definition larger than 1.

As shown in Figure 4.2, Method 1 improves on Denton for the non-monotonic series. For monotonic series, the results are slightly worse for trade and employment and somewhat better for retail. Method 2 outperforms Denton for the non-monotonic trade and retail series. The most striking result is the lack of improvement for employment. An explanation follows below. From Table 4.3 one can see that the growth rates of monthly and quarterly data are relatively well consistent for employment. Therefore, sequential Denton does not produce serious step problems between sequentially benchmarked series. Method 2 still achieves better movement preservation around the boundaries of the estimation intervals, but the improvement is relatively small and does not compensate for the lower degree of movement preservation within benchmarked series. Table 4.4 further illustrates this. This table compares movement preservation between and within sequentially benchmarked series.

Table 4.4 Movement preservation; employment; non-monotonic series;

	Seq. Denton	Method 1	Method 2
Total Movement preservation	9.78	9.60	10.00
- within seq. benchmarked series	8.96	8.96	9.64
- between seq. benchmarked series	0.82	0.63	0.36

Movement preservation is based on the objective function (4.5). All values are $*10^{-4}$

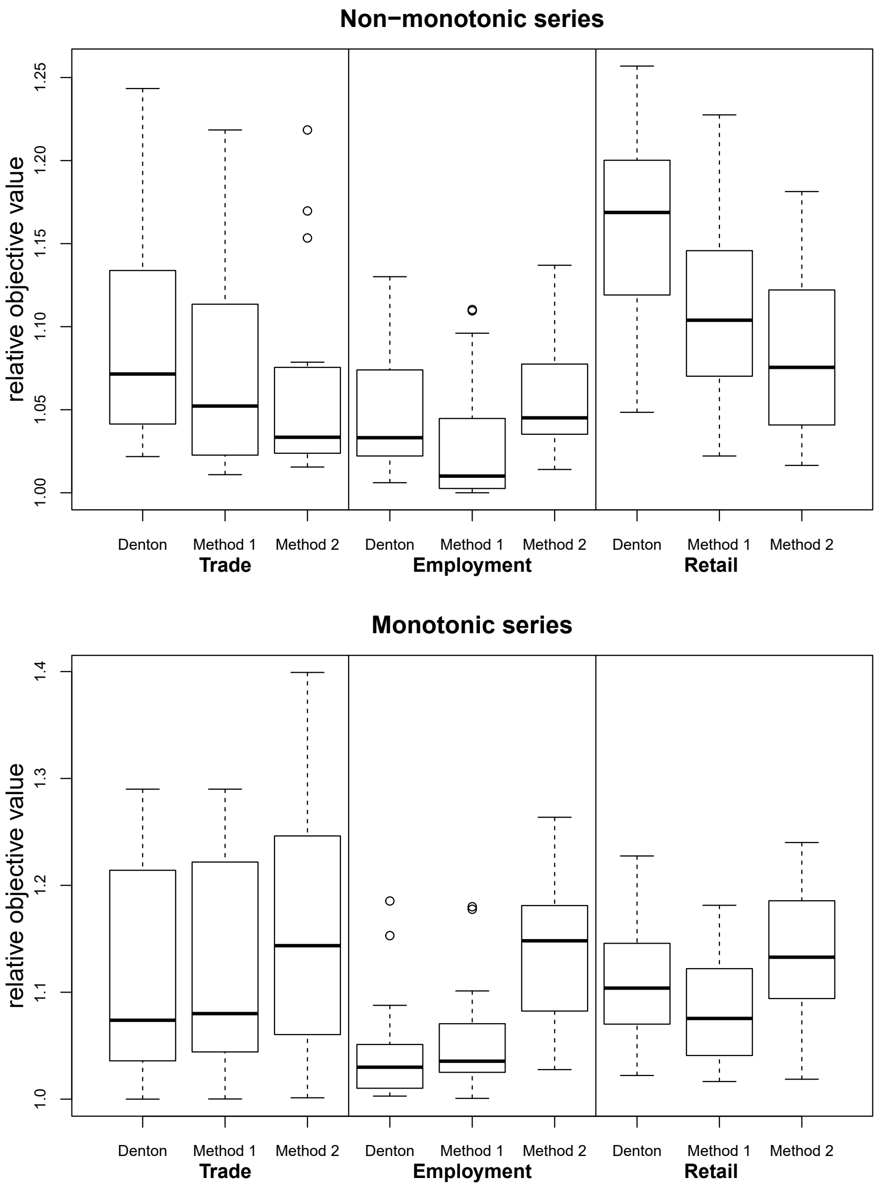

Figure 4.2 Relative objective values of three sequential methods compared to optimal simultaneous Denton

4.5.3 Results: Relative distance benchmarked values

A second evaluation criterion compares the benchmarked values of a sequential method with those of simultaneous Denton-Cholette. A relative distance criterion is used, as expressed in (4.10). In this formula $x_{\text{seq, t}}$ stands for the benchmarked value for period t from a sequential method and $x_{\text{sim,t}}$ is the result from simultaneous Denton-Cholette for the same series.

$$\text{Rel. Difference} := \frac{100}{n}\sum_{t=1}^{n}\frac{|x_{seq,t}-x_{sim,t}|}{x_{sim,t}}. \tag{4.10}$$

Figure 4.3 shows a boxplot of the relative differences for all ‘non-monotonic’ series. Usually, the newly proposed sequential methods outperform the existing Denton method, as the benchmarked values are closer to the sequentially benchmarked ones. However, for the employment data the results of Method 2 are not always better. An explanation for this has already been given in previous subsection.

4

4.5.4 Summary of results

To summarize the previous results, Method 1 improves on Denton for most non-monotonic series. Often, Method 2 produces even better results, but only if temporal discrepancies are sufficiently large. Theoretically, Method 2 puts disproportional focus on the preservation of the last values. The preservation of the last value is modelled as a hard constraint, whereas the preservation of the first value and the short-term movements can be considered ‘soft constraints’, as some deviation from the target values is allowed. This explains why Method 2 works well for problems in which Denton produces relatively large steps between sequentially benchmarked series, but performs worse in the absence of large step problems. As Method 1 can be considered a more balanced method, the remainder of this chapter deals with Method 1 only.

4.6 GENERIC SOLUTIONS

Section 4.4 introduced two sequential benchmarking methods. These methods are especially useful for series whose benchmarking corrections evolve non-monotonically between estimation intervals. In practice, it cannot be known whether this condition is fulfilled. This section proposes to use historical data to choose a suitable benchmarking algorithm. In this way, a practically feasible solution is created that can be applied to any data set, conditional on the availability of historical data.

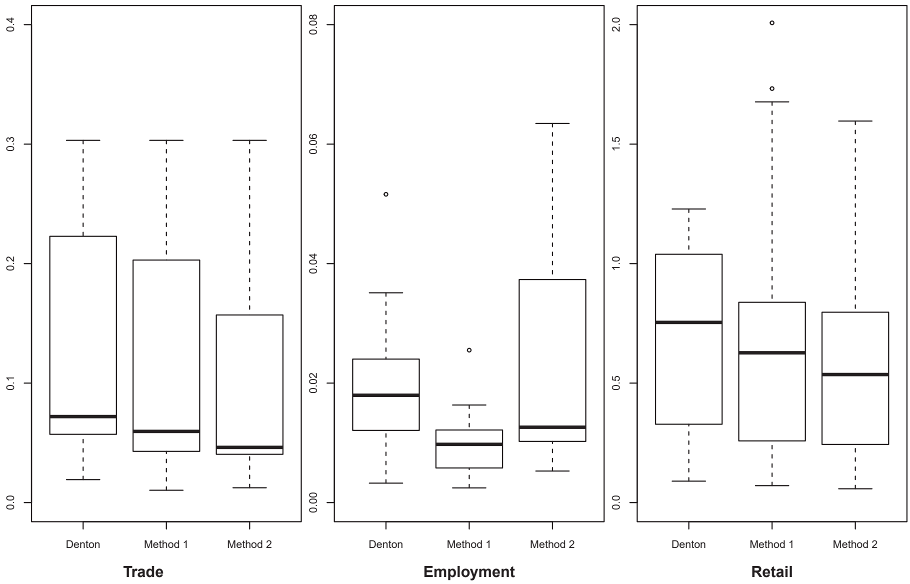

Figure 4.3 Relative difference benchmarked values; non-monotonic series

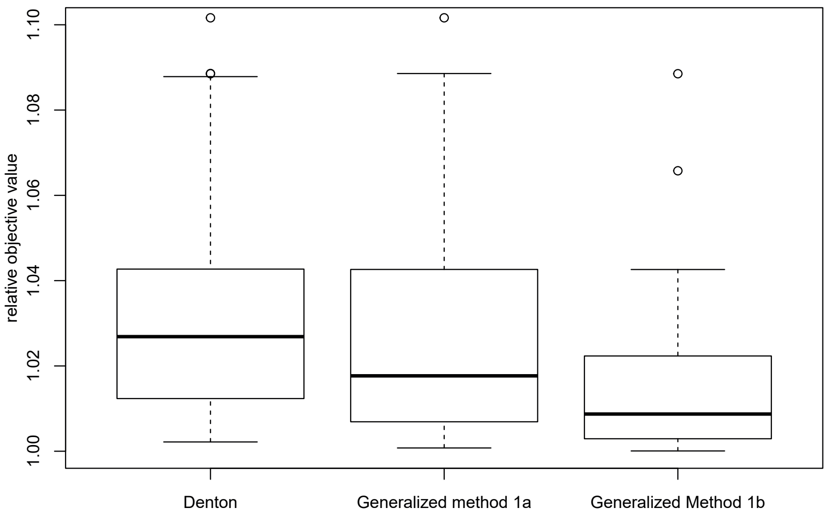

Figure 4.4 Relative movement preservation sequential benchmarking, as in Section 4.5.2

4.6.1 Methods

This subsection proposes two generalisations of Method 1 in Section 4.4, that are referred to as Method 1a and Method 1b hereafter.

Generalized method 1a

This generalized method means that a selection criterion is applied to judge the suitability of Method 1. Time series that meet the criterion are benchmarked with Method 1. All other series are processed by Denton PFD. The selection criterion uses historical data. Similarly to Section 4.5.1, it selects series for which LF benchmarking corrections change non-monotonically between estimation intervals for more than half of the cases. That is, for which $BI_{N+1} \geq BI_{N+1} \geq BI_{N-1}$ or $BI_{N+1} \leq BI_{N+1} \leq BI_{N-1}$for the majority of past estimation intervals.

Generalized method 1b

The second generalized method extends Method 1's objective function in (4.7) and provides a criterion to select series that are eligible for the extended method. The extended method explicitly uses past corrections for the estimation of current interval. As stated by Di Fonzo and Marini (2012): "when a systematic pattern arises from the annual series of the BI ratio, especially in the latest years, the user should take advantage of this information and try to exploit it".

The objective function of the extended method is stated as follows

$$f^{Method\ 1b}(\boldsymbol{x}) = \left(\frac{x_1}{p_1} - 1\right)^2 + \left(\frac{x_n}{t_n^*} - 1\right)^2 + \sum_{t=2}^{n}\left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}}\right)^2 \tag{4.11}$$

where $t_n^* = p_n c_n^* BI_N$ and c_n^* is a correction factor.

Method 1 in Section 4.4 emerges as the special case in which c_n is set to one. Generalized method 1b aims to use $c_n = BI_{N+1} \ / \ BI_{\mathrm{N}}$. By this choice, the benchmarking adjustment for n is moved into the direction of the future benchmarking correction BI_{N+1}. A value for c_n is estimated from historical data. It is the median value of $BI_{N+1} \ / \ BI_N$, taken over previous estimation intervals. That is,

$$c_n = median_{s=1,..,S}\left(\frac{BI_{N+1-sN}}{BI_{N-sN}}\right)$$

where S stands for the number of previous estimation intervals. The formula is chosen on an ad-hoc basis. Alternative expressions can be developed in the future.

The generalised method 1b can be expected to improve on Denton, if the differences between the actual and the Denton assumed values for BI_{N+1} are clear and consistent in time. To select such series, an ad-hoc criterion is derived below.

Denton assumes that the trend of the reconciliation adjustments at the end of a previous estimation interval continues at the beginning of the following interval. That is, $BI_{N+1} \approx BI_N(BI_N/BI_{N-1})$. Conversely, Denton cannot be expected to perform very well if ΔBI_{N+1} structurally differs from ΔBI_N, where $\Delta BI_N = BI_{N+1}/BI_N$. Therefore, the selection criterion aims to select the series with a structural difference between ΔBI_N and ΔBI_{N+1}.

The criterion assumes that sets of historical values are available for ΔBI_N and ΔBI_{N+1}, denoted by $\Delta \mathrm{BI}_{N\text{-}sN}$ and $\Delta BI_{N+1\text{-}sN}$, where $s = 1, \ldots, S$ and S stands for the number of past estimation intervals. Quartiles are computed for these two sets of observations. It is said that there is a structural difference between $\Delta BI_N$and ΔBI_{N+1} if

$$\text{First Quartile}_{s=1,\ldots,S}\ (\Delta \mathrm{BI}_{N\text{-}sN}) > \text{Third Quartile}_{s=1,\ldots,S}(\Delta BI_{N+1\text{-}sN}) \text{ or}$$
$$\text{Third Quartile}_{s=1,\ldots,S}\ (\Delta \mathrm{BI}_{N\text{-}sN}) < \text{First Quartile}_{s=1,\ldots,S}(\Delta BI_{N+1\text{-}sN})$$

In the first case, ΔBI_N is often higher than ΔBI_{N+1}, in the second case it is the other way around.

All series that satisfy the above-mentioned criterion are benchmarked according to the objective function in (4.11). Denton PFD is applied to all other series. Method 1b may perform better than Method 1a because of its higher flexibility. A risk of Method 1b is however that results may not be that good if there is an influential change in the pattern of benchmarking adjustments.

4.6.2 Application

The two generalized methods have been applied to the employment data, as described in Section 4.5.1. Data for 2011-2016 were benchmarked, assuming that the results for 2004-2010 are known. On the basis of the previously mentioned criteria and the 2004-2010 data 18 out of 53 series have been selected for Method 1a and 49 series were considered suitable for Method 1b.

As shown in Figures 4.4 and 4.5, the two generalised methods lead to better results than Denton. The improvement is the largest for Generalised method 1b. This points out that it can be useful to predict future reconciliation adjustments based on past adjustments

4.7 CONCLUSIONS

This chapter considers a sequential benchmarking problem, in which HF series are sequentially benchmarked to LF totals. The problem is relevant for a 'production setting', in which previously published results are not allowed to be revised after some time. It was demonstrated that a popular Denton method does not always yield satisfactory results.

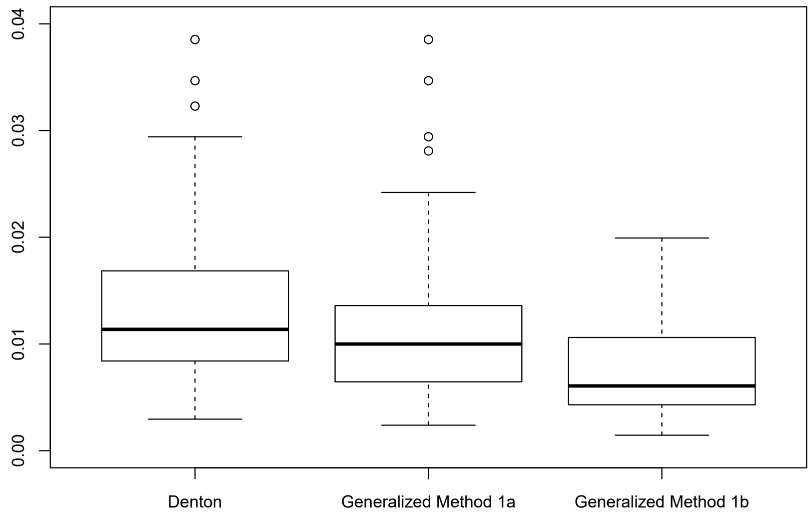

Figure 4.5 Relative difference benchmarked values, as in Section 4.5.3

Abrupt changes in benchmarking corrections can occur, especially around the boundaries of the estimation intervals and corrections of preliminary values can be unnecessarily large. This chapter proposed two new methods for non-monotonic time series and illustrates their application with empirical results. Section 4.6 further enhances and generalizes the first method with two variants called generalized method 1a and generalised method 1b. Method 1b assumes that historical benchmarking corrections reoccur at later points in time. Method 1a is more neutral, in the sense that it does not explicitly exploit historical adjustment patterns when estimating the current period. Compared to Denton PFD, the newly proposed methods produce smoother benchmarking corrections at the beginning and end of the estimation intervals. Movement preservation within estimation intervals might however be worse. Therefore, the newly proposed methods can only be advised for specific classes of applications.

Section 4.6 proposed selection criteria to decide on the appropriateness of the two new methods for a time series at hand. The proposed criteria rely on historical benchmarking corrections. The criteria have been developed in an ad-hoc manner. The proposed criteria might be refined in the future. Alternatively to formal criteria, expert knowledge can be used to decide on the estimation method. If there is any reason to assume that the benchmarking adjustments are structurally different at the beginning and at the end of each estimation interval, the newly proposed methods can be expected to be useful. This can happen for instance if the LF and HF series exhibit different seasonal patterns a situation that is most often observed for month-to-quarter benchmarking.

The empirical applications in Sections 4.5 and 4.6 demonstrate that the new methods notably improve on Denton PFD's results, with respect to movement preservation and the closeness of benchmarked values to simultaneous benchmarked results.

As a final remark, it was assumed in this chapter that when benchmarking the current interval, no data are available for the beginning of the sequential estimation interval. If however LF and HF data are available for at least one LF period after the end of the current interval, it would be advisable to use these data when benchmarking the current series. Since benchmarked results for the periods beyond the current estimation interval are of no interest to a practitioner, these results are discarded. The merit of the solution is however that the reconciliation adjustments of the current interval move already into the direction of the future adjustments, thus reducing the future correction in the following estimation interval.

5. Divide-and-Conquer solutions for estimating large consistent table sets[4,5]

Summary. When several frequency tables need to be produced from multiple data sources, there is a risk of numerically inconsistent results. This means that different estimates are produced for the same cells or marginal totals in multiple tables. As inconsistencies of this kind are often not tolerated, there is a clear need for compilation methods for achieving numerically consistent output. Statistics Netherlands developed a Repeated Weighing (RW) method for this purpose. The scope of applicability of this method is however limited by several known estimation problems. This chapter presents two new Divide-and-Conquer (D&C) methods, based on quadratic programming (QP) that avoid many of the problems experienced with RW.

4 This chapter has been published as: Daalmans J.A. (2018) Divide-and-Conquer solutions for estimating large consistent table sets. *Statistical Journal of the IAOS*, 34, 223-233.

5 This chapter is based on work carried out as part of the Eurostat project "Improvement of the use of administrative sources" (ESS.VIP ADMIN WP6 Pilot studies and applications). The action has received EU funding under the grant agreement 07112.2015.002-2015.353. The chapter reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains. The author thanks Ton de Waal, Tommaso di Fonzo, Nino Mushkudiani, Eric Schulte Nordholt, Frank Linder, Arnout van Delden and Reinier Bikker for useful suggestions.

5.1 INTRODUCTION

Statistical outputs are often interconnected. Different tables may share common cells or marginal totals. In such cases numerically consistency is often required, i.e. that the same values are published for common outputs. However, due to different data sources and compilation methods, numerically consistency is often not automatically achieved. Hence, there is a clear need for methods for achieving numerically consistent output.

An important example of a multiple table statistical output in the Netherlands is the Dutch Population and Housing Census. For the census, dozens of detailed contingency tables need to be produced with many overlapping variables. Numerically consistent results are required by the European Census acts and a number of implementing regulations (European Commission, 2008). In a traditional census, based on a complete enumeration of the population, consistency is automatically present. Statistics Netherlands belongs to a minority of countries that conducts a virtual census. In a virtual census estimates are produced from already available data that are not primary collected for the census. The Dutch virtual census is for a large part based on integral information from administrative sources. For a few variables not covered by integral data sources, supplemental sample survey information is used. Because of incomplete data, census compilation relies on estimation. Due to the different data sources that are used numerically inconsistent results would be inevitable if standard estimation techniques were applied (De Waal, 2015 and 2016).

To prevent inconsistency, Statistics Netherlands developed a method called "Repeated Weighting" (RW), a method that was applied to the 2001 and 2011 Censuses, see e.g.), e.g. Renssen and Nieuwenbroek (1997), Nieuwenbroek *et al.* (2000), Houbiers *et al.* (2003) and Knottnerus and Van Duin (2006). In RW the problem of consistently estimating a number of contingency tables with overlapping variables is simplified by splitting the problem into dependent sub problems. In each of these sub problems a single table is estimated. Thus, a sequential estimation process is obtained.

The implementation of RW is however not without its problems, see e.g. Houbiers *et al.* (2003), Daalmans (2015) and Subsection 5.2.4 below. In particular, there are problems that are directly related to the sequential approach. Most importantly, RW does not always succeed in estimating a consistent table set, even when it is clear that such a table set exists. After a certain number of tables have been estimated, it may become impossible to estimate a new one consistently with all previously estimated ones. This problem seriously limits future application possibilities of repeated weighting. For the Dutch 2011 Census several ad-hoc solutions were applied, designed after long trial-and-error. For any future application, it is however not guaranteed that numerically consistent estimates can be produced. Therefore, there is a clear need for extending methodology.

This chapter presents two new ‘Divide-and-Conquer algorithms’ based on Quadratic Programming (QP). The algorithms break down the problem of consistent estimation of a set of contingency tables into sub problems that can be independently estimated, rather than the dependent parts that are obtained in RW. Thus, the estimation problems, as experienced with RW, are avoided.

In Section 5.2, we describe the RW method. Section 5.3 presents an alternative quadratic programming (QP) formulation for this problem. Section 5.4 gives a simultaneous weighting approach, which is the basis for the two new Divide-and-Conquer methods that are introduced in Section 5.5. Results of a practical application are given in Section 5.6.

5.2 REPEATED WEIGHTING

5.2.1 Prerequisites

Although RW can be applied to contingency and continuous data, this chapter deals with contingency tables only.

I assume that multiple prescribed tables need to be produced with overlapping variables. If there were no overlapping variables, it would not be any challenge to produce numerically consistent estimates.

Further, it is assumed that the target populations are the same for each table. This means for example that all tables necessarily have to add up to the same grand-total.

All data sources relate to the same target population. There is no under- or overcoverage: the target population of the data sources coincides with the target population of the tables to be produced.

Further, for each target table a predetermined data set has to be available from which that table is compiled.

Two types of data sets will be distinguished: data sets that cover the entire target population and data sets that cover a subset of that population. As the first type is often obtained from (administrative) registers and the latter type from statistical sample surveys, these data sets will be called registers and sample surveys from now on.

It is assumed that all register-based data sets are already consistent at the beginning of RW. That means that all common units in different data sets have the same values for common variables. Subsection 5.2.2 explains why this assumption is important. In practice, this assumption often means that a so-called micro integration process has to be applied prior to repeated weighting (Bakker *et al.*, 2014).

For sample survey data sets it is required that weights are available for each unit that are meant to be used to draw inferences for a population. To obtain weights for sample surveys, one usually starts with the sample weight, i.e. the inverse of the probability of selecting a unit in the sample. Often, these sample weights are adjusted to take selectivity or non-response into account. Resulting weights will be called starting weights, as these are weights that are available at the beginning of repeated weighting.

5.2.2 Non-technical description

The compilation method of a single table depends on the type of the underlying data set.

Tables that are derived from a register are simply produced by counting from that register. This means that for each cell in the table, it is counted how much the corresponding categories occur (e.g. 28 year old males). There is no estimation involved, because registers are supposed to cover the entire target population. The fact that register-based data are not adjusted explains why registers need to be already consistent at the beginning.

Below we focus on tables that are derived from a sample survey. These tables have to be consistently estimated. This basically means two things: common marginal totals in different tables have to be identically estimated and all marginal totals for which known register values exist have to be estimated consistently with those register values.

In the RW-approach consistent estimation of a table set is simplified by estimating tables in sequence. The main idea is that each table is estimated consistently with all previously estimated tables. When estimating a new table, it is determined first which marginal totals the table has in common with all registers and previously estimated tables. Then, the table is estimated, such that:

1) Marginal totals that have already been estimated before are kept fixed to their previously estimated values;
2) Marginal totals that are known from a register are fixed to their known value.

To illustrate this idea, we consider an example in which two tables are estimated:

Table 1: age × sex × educational attainment;

Table 2: age × geographic area × educational attainment.

A register is available that contains age, sex and geographic area. Educational attainment is available from a sample survey. Because educational attainment appears in Table 1 and 2, both tables need to be estimated from that sample survey. To achieve consistency, Table 1 has to be estimated, such that its marginal totals age × sex aligns with the known population totals from the register. For Table 2 it needs to be imposed that the marginal total age × geographic area complies with the known population totals from the register and that the marginal total age × educational attainment is estimated the same as in Table 1.

Each table is estimated by means of the generalised regression estimator (GREG) Särndal *et al.* (1992), an estimator that belongs to the class of calibration estimators Deville *et al.* (1992). Thus, repeated weighting comes down to a repeated application of the GREG-estimator.

5

5.2.3 Technical description

In this subsection, repeated weighting is described in a more formal way. Below we will explain how a single table is estimated from a sample survey.

Aim of the repeated weighting estimator (RW-estimator) is to estimate the P cells of a frequency table $\mathbf{Y}_1, \ldots, \mathbf{Y}_P$. We will use vector notation to express the elements of a table. The estimates are made from a sample survey, of which initial, strictly positive weights w_i are available for all n records. Each record in the microdata contributes to exactly one of the cells of a table. A dichotomous variable y_{ip} is used, which is one if record i contributes to cell p and zero otherwise.

A simple population estimator is given by

$$\hat{\boldsymbol{t}}_y^w = \sum_{i=1}^{n} w_i \boldsymbol{y}_i,$$

where y_i is a P-vector, containing the elements y_{ip} for $p = 1, \ldots, P$. The estimator $\hat{\boldsymbol{t}}_y^w$ is obtained by aggregation of starting weights of the data set used for estimation.

The so-called initial table estimate $\hat{\boldsymbol{t}}_y^w$ is independent of all other tables and registers and is not necessarily consistent with other tables. To realize consistency, a population estimate needs to be calibrated on all marginal totals that the table has in common with all registers and with all previously estimated tables. These marginal totals are denoted by the J-vector $\boldsymbol{r}$.

There is a relationship between the cells of a table and its marginal totals: a marginal total is a collapsed table that is obtained by summing along one or more dimensions. Each cell contributes to a specific marginal total or it does not. The relation between the P cells and the J marginal totals is expressed in an $(J \times P)$- aggregation matrix L. An element l_{jp} is 1 if cell p of the target table contributes to marginal total j and zero otherwise.

A table estimate $\hat{\boldsymbol{t}}_y$ is consistent if it satisfies

$$\mathbf{L}\hat{\boldsymbol{t}}_y = \boldsymbol{r}. \tag{5.1}$$

Usually, initial estimates $\hat{\boldsymbol{t}}_y^w$ do not satisfy (5.1), otherwise no adjustment would be necessary.

Therefore, our aim is to find a table estimate $\hat{\boldsymbol{t}}_y^*$ that is in some sense close to $\hat{\boldsymbol{t}}_y^w$ and that satisfies all consistency constraints. The well-established technique of least-square adjustment can be applied to find such an adjusted estimate. In this approach, a consistent table estimate $\hat{\boldsymbol{t}}_y^*$ is obtained as a solution of the following minimization problem

$$\min_{\hat{\boldsymbol{t}}_y^*} \left(\hat{\boldsymbol{t}}_y^* - \hat{\boldsymbol{t}}_y^w\right)' \mathbf{W}^{-1} \left(\hat{\boldsymbol{t}}_y^* - \hat{\boldsymbol{t}}_y^w\right),$$
$$\text{such that: } \mathbf{L}\hat{\boldsymbol{t}}_y^* = \boldsymbol{r}. \tag{5.2}$$

where **W** is a symmetric, non-singular weight matrix.

Despite that several alternative methods can be applied as well, e.g. Deville *et al.* (1992) and Little and Wu (1991), the Generalised Least Squares (GLS) problem in (5.2) has a long and solid tradition in official statistics.

A closed-form expression for the solution of the problem in (5.2) can be obtained by the Lagrange Multiplier method (see e.g. Mushkudiani *et al.* (2015)). This expression is given by

$$\hat{\boldsymbol{t}}_y^{opt} \;=\; \hat{\boldsymbol{t}}_y^{w} + \mathbf{WL}'\,(\mathbf{LWL}')^{-1}\left(\boldsymbol{r} - \mathbf{L}\,\hat{\boldsymbol{t}}_y^{w}\right). \tag{5.3}$$

The GREG-estimator is obtained as special case of (5.3) in which **W** is set to $\widehat{\mathbf{T}}$, where $\widehat{\mathbf{T}}$ = diag$\left(\hat{\boldsymbol{t}}_y^{w}\right)$, a diagonal matrix with the entries of $\hat{\boldsymbol{t}}_y^{w}$ along its diagonal (Deville *et al.*, 1992) . Thus, we obtain the following expression for the RW-estimator.

$$\hat{\boldsymbol{t}}_y^{RW} \;=\; \hat{\boldsymbol{t}}_y^{w} + \widehat{\mathbf{T}}\mathbf{L}'\left(\mathbf{L}\widehat{\mathbf{T}}\mathbf{L}'\right)^{-1}\left(\boldsymbol{r} - \mathbf{L}\,\hat{\boldsymbol{t}}_y^{w}\right), \tag{5.4}$$

In writing (5.4), it is assumed that the inverse of square matrix $\mathbf{L}\widehat{\mathbf{T}}\mathbf{L}'$ is properly defined. In practice, this is however not always true. When the constraint set in (5.1) contains any redundancies, i.e. constraints that are implied by other constraints, $\mathbf{L}\widehat{\mathbf{T}}\mathbf{L}'$ will be singular. In that case, it may still be possible to apply (5.4) by using a generalised inverse e.g. Ben-Israel and Greville (2003).

As an alternative to minimizing adjustment at cell level, the RW solution can also be obtained by minimal adjustment of underlying weights. In Deville *et al.* (1992) it is shown that a set of adjusted weights w_{ip}^{*} can be derived, such that the RW estimate $\hat{\boldsymbol{t}}_y^{RW}$ can be obtained by weighting the underlying micro data. That is, such that:

$$\left(\hat{\boldsymbol{t}}_y^{RW}\right)_p \;=\; \sum_{i=1}^{n} w_{ip}^{*} y_{ip}. \tag{5.5}$$

For data sets that underlie estimates for multiple tables, adjusted weights are usually different for each table.

From the expression for the RW-estimator in (5.4), it follows that initial cell estimates of zero remain zero, since the relevant rows in $\widehat{\mathbf{T}}\mathbf{L}'\left(\mathbf{L}\widehat{\mathbf{T}}\mathbf{L}'\right)^{-1}$ contain zeros only. However, in presence of zero-valued initial estimates, the so-called empty cell problem may occur. This happens if there is a constraint imposing a sum of variables that each has a zero initial estimate to align with a nonzero value in **r**. Because zero values cannot be adjusted, achieving consistency is impossible. The RW estimator in (5.4) is undefined because $\mathbf{L}\widehat{\mathbf{T}}\mathbf{L}'$ includes an all zeroes row. Consequently, the originally proposed RW-method cannot be applied if the empty cell problem occurs.

Besides reconciled table estimates, RW also provides means to estimate precision of these estimates. Variances of table estimates can be estimated, see Houbiers *et al.* (2003) for mathematical expressions.

5.2.4 Problems with repeated weighting

Below we summarise complications of RW. Problems that are inherent to the sequential way of estimation are described first, then other complications are given.

Problem 1. Impossibility of consistent estimation

A first problem of RW is that, after a number of tables have been estimated, it may become impossible to estimate a new one. Earlier estimated tables impose certain consistency constraints on a new table, which reduces the degree of freedom for the estimation of that new table. When a number of tables have already been estimated it may become impossible to satisfy all consistency constraints at the same time. The problem is also known in literature (Cox, 2003), for the estimation of multi-dimensional tables with known marginal totals.

Example

Suppose one wants to estimate the table: country of citizenship × industry of economic activity × educational attainment. Citizenship and industry are observed in a register, educational attainment comes from a survey. According to the register there are: 10 persons from Oceania and 51 persons working in the mining industry. The combination Oceania and mining industry is observed for four persons. The marginal totals as derived from previously estimated tables are shown in Tables 5.1 and 5.2.

Table 5.1. Marginal total 1

Citizenship	Education	Count
Oceania	Low	1
Oceania	High	9

Table 5.2. Marginal total 2.

Industry	Education	Count
Mining	Low	49
Mining	High	2

By combining both tables, it can be seen that the combination Oceania & mining industry can occur three times at most; there cannot be more than two highly educated people and one lowly educated person. This contradicts results from the register that states that there are four “mining” persons from Oceania. The problem occurs because the known population counts for the combination of citizenship and industry are not taken into account in the previously estimated tables.

Problem 2. Suboptimal solution

In the RW-approach the problem of estimating a set of coherent tables is split into a number of sub problems, in each of which one table is estimated. Because of the sequential approach, a suboptimal solution may be obtained that deviates more from the data sources than necessary.

Problem 3. Order dependency

The order of estimation of the different tables matters for the outcomes. Besides that ambiguous results are not desirable as such, it can be expected that there is a relationship between the quality of the RW-estimates and the order of estimation, as tables that are estimated at the beginning of the process do not have to satisfy as many consistency constraints as tables that are estimated later in the process.

In addition to the aforementioned problems, there are also some other problems that are not directly caused by sequential estimation.

A first problem is that although RW achieves consistency between estimates for the same variable in different tables, the method does not support consistency rules between different variables (so-called 'edit rules'). An example of such a rule is that the number of people who have never resided abroad cannot exceed the number of people born in the country concerned.

A second complication is that RW may yield negative cell estimates. In many practical applications, such as the Dutch Census, negative values are however not allowed.

A third complication is the previously mentioned empty cell problem. As mentioned in Subsection 5.2.3, this problem occurs when estimates have to be made without underlying data. It is caused by sampling effects, i.e. characteristics that are known to exist in the population that are not covered by a sample survey used for estimation. The empty cell problem can be tackled by the epsilon method: a technical solution proposed by Houbiers (2013) based on the pseudo-Bayes estimator for log-linear analysis (Bishop *et al.*,1975). The epsilon method means that zero-valued estimates in an initial table are replaced by small, artificial, non-zero "ghost" values, which were set to one for all empty cells in the 2011 Census tables. In other words, it was assumed a priori that each empty cell is populated by one fictitious person.

5.3 REPEATED WEIGHTING AS A QP PROBLEM

This section demonstrates that the consistent estimation problem can alternatively be solved by available techniques from Operations Research (OR). The repeated weighting estimator in (5.4) can be obtained as a solution of the following quadratic programming problem (QP).

$$\begin{aligned}
&\underset{\hat{\boldsymbol{t}}_y^*}{min} \sum_{i:\,(\hat{\boldsymbol{t}}_y^w)_i>0} \frac{1}{(\hat{\boldsymbol{t}}_y^w)_i}\left((\hat{\boldsymbol{t}}_y^*)_i - (\hat{\boldsymbol{t}}_y^w)_i\right)^2,\\
&\text{such that:}\\
&\mathbf{L}\hat{\boldsymbol{t}}_y^* = \boldsymbol{r},\\
&(\hat{\boldsymbol{t}}_y^*)_i = 0 \quad \text{for } i \text{ with } (\hat{\boldsymbol{t}}_y^w)_i = 0.
\end{aligned} \qquad (5.6)$$

The objective function minimizes squared differences between reconciled and initial estimates. The constraints are the same as in RW. The last mentioned type of constraint ensures that zero-valued estimates are not adjusted.

The main advantage of the QP-approach is its computational efficiency. Unlike the closed-form expression of the RW estimator (5.4), Operations Research methods do not rely on matrix inversion. Therefore, very efficient solution methods are available (e.g. Nocedal and Wright (2006)). Operations Research methods are available in efficient software implementations ('solvers'), that are able to deal with large problems. Examples of well-known commercial solvers are Xpress (Dash Optimization, 2017), Gurobi (Gurobi, 2016) and Cplex (IBM, 2015).

At Statistics Netherlands, mathematical optimization methods are applied for National Accounts balancing, Bikker *et al.* (2013), an application that requires solving a quadratic optimization problem of approximately 500,000 variables.

A second advantage of the QP-approach is that it can still be used in case of redundant constraints. Contrary to the WLS-approach, there is no need to remove redundant constraints, or to apply sophisticated techniques like generalised inverses.

A third advantage is that QP can be more easily generalised than WLS to include additional requirements. Inequality constraints can be included in the model to take account of non-negativity requirements and edit rules (see Subsection 5.2.4). The empty cell problem can be dealt with by the following slight modification of the objective function

$$\begin{aligned}
&\underset{\hat{\boldsymbol{t}}_y}{\min} \sum_{i=1}^{P} \frac{1}{(\hat{\boldsymbol{t}}_y^{w^*})_i}\left((\hat{\boldsymbol{t}}_y^*)_i - (\hat{\boldsymbol{t}}_y^w)_i\right)^2,\\
&\text{such that:}\\
&\mathbf{L}\hat{\boldsymbol{t}}_y^* = \boldsymbol{r}.
\end{aligned} \qquad (5.7)$$

where $\hat{\boldsymbol{t}}_y^{w^*} = pmax\left(\hat{\boldsymbol{t}}_y^w, 1\right)$ and *pmax* stands for pairwise maximum. The solution in (5.7) is less radical than replacing each initial zero estimate with one, the solution that was applied for the 2011 Dutch census. The objective function in (5.6) is a weighted sum of squared differences. The weights are changed in (5.7), but the quadratic terms are the same as in (5.6).

Disadvantages of the QP-approach are that the method does not provide means to derive corrected weights and to estimate variances of reconciled tables. However, because of the equivalence of the QP and the WLS formulation of the problem, it follows that, although corrected weights are not obtained in a solution of a QP-problem, these weights do exist from a theoretical point of view.

5.4 SIMULTANEOUS APPROACH

In this section we argue that the three problems mentioned in Subsection 5.2.4 ("Impossibility of consistent estimation", "Suboptimal solution" and "Order dependency") that are inherent to the sequential way of estimation can be circumvented in an approach in which all tables are estimated simultaneously. The QP-model in (5.6) can be easily generalized for the consistent estimation of a table set. That is, a consistent table set can be obtained as a solution to the following problem

$$\begin{aligned}
&\min_{\hat{\boldsymbol{t}}^{SW}} \sum_{i:\,(\hat{\boldsymbol{t}}^{w})_i>0} \frac{1}{(\hat{\boldsymbol{t}}^{w})_i}\left((\hat{\boldsymbol{t}}^{SW})_i - (\hat{\boldsymbol{t}}^{w})_i\right)^2, \\
&\text{such that:} \\
&\mathbf{L}\hat{\boldsymbol{t}}^{SW} = \boldsymbol{r}. \\
&(\hat{\boldsymbol{t}}^{SW})_i = 0, \text{ for } i \text{ with } (\hat{\boldsymbol{t}}^{w}_y)_i = 0.
\end{aligned} \tag{5.8}$$

5

In this formulation $\hat{\boldsymbol{t}}^{SW} = \left(\hat{\boldsymbol{t}}^{SW}_{1,\ldots,}\hat{\boldsymbol{t}}^{SW}_{N}\right)'$ is a vector containing estimates for the cells of all N tables, similarly $\hat{\boldsymbol{t}}^{w} = \left(\hat{\boldsymbol{t}}^{w}_{1,\ldots,}\hat{\boldsymbol{t}}^{w}_{N}\right)'$, a vector of initial estimates. The subscript SW stands for simultaneous weighting, as opposed to RW, which stands for repeated weighting.

The objective function minimises a weighted sum of squared differences between initial and reconciled cell estimates for all tables. The constraints impose marginal totals of estimated tables to be consistent with known population totals from registers and estimated tables to be mutually consistent. The former means that for each table all marginal totals with known register totals are consistently estimated with those register totals. The latter means that for each pair of two distinct tables all common marginal totals have the same estimated counts. These constraints impose a sum of cells in one table to be equal to a sum of cells in another table, where the value of that sum is not known in advance. For comparison, in RW, marginal totals of one table need to have the same value as known marginal totals from earlier estimated tables. Analogous to the RW-model in (5.6), the SW-model in (5.8) can be easily extended to take account of additional requirements, like non-negativity of estimated cell values, edit rules and the empty cell problem.

It can be easily seen that Problems 1, 2 and 3 in Subsection 5.2.4 do not occur if all tables are estimated simultaneously. Furthermore, from a practical point it is more attractive to solve one problem rather than several problems.

A SW-approach may however not always be computationally feasible. A large estimation problem needs to be solved consisting of many variables and constraints. The capability of solving such large problems may still be limited by computer memory size, even for modern computers. We therefore focus on ways of splitting the problem up into a number of smaller sub problems that can be preferably independently solved.

5.5 DIVIDE-AND-CONQUER ALGORITHMS

In this section two so-called Divide-and-Conquer (D&C) algorithms are presented for estimating a set of coherent frequency tables. These algorithms recursively break down a problem into sub problems that can each be more easily solved than the original problem. The solution of the original problem is obtained by combining the sub problem solutions.

5.5.1 Splitting by common variables

The main idea of our first algorithm is that an estimation problem, with one or more common register variable(s) can be split into a number of independent sub problems, based on the categories of these register variable(s). For example, if sex were included in all tables of a table set, a table set can be split into two independent sets: one for men and one for women.

In practice, it is often not the case that a table set has one or more common register variables in each table. Common variables can however always be created by adding variables to tables, provided that a data source is available from which the resulting, extended tables can be estimated. In our previous example, all tables that do not include sex can be extended by adding this variable to the table. In this way, the level of detail increases, meaning that more cells need to be estimated as in the original problem, which may come along with a loss of precision at the required level of publication. However, at the same time, the possibility is created of splitting a problem into independent sub problems. Since all 'added' variables are used to split the problem, one can easily understand that the number of cells in each of these sub problems cannot exceed the total number of cells of the original problem.

For any practical application the question arises which variable(s) should be chosen as "splitting" variable(s). Preferably, this/ these should be variable(s) that appear in most tables, e.g. sex and age in the Dutch 2011 Census, as this choice leads to the smallest total number of cells to be estimated.

The approach is especially useful for a table set with many common variables, because in that case the number of added cells remains relatively limited.

The proposed algorithm has the advantage over Repeated Weighting that the sub problems that are created can be solved independently. For this reason there are no problems with "impossibility of estimation" (Problem 1 in Subsection 5.2.4) and "order-dependency of estimation" (Problem 3 in Subsection 5.2.4). Problem 2 "Suboptimal solution" is not necessarily solved. This depends on the need of adding additional variables to create common variables. If a table set contains common register variables in each table and the estimation problem is split using these common variables, an optimal solution is obtained. However, if common variables are created by adding variables to tables, extended tables are obtained, for which the optimal estimates do not necessarily comply with the optimal estimates for the original tables.

5.5.2 Aggregation and disaggregation

A second divide-and-conquer algorithm consists of creating sub problems by aggregation of one or multiple variable categories. In the first stage, categories are aggregated (e.g. estimating 'educational attainment' according to two categories rather than the required eight). In a second stage, table estimates that include the aggregation variable(s) are further specified according to the required definition of categories.

Since the disaggregation into required categories can be carried out independently for each aggregated category, a set of independent estimation problems is obtained in the second stage.

For example, suppose that we need to estimate educational attainment, according to 8 categories: 1,…,8. Two aggregated categories I and II are defined; I comprises the original categories 1,…,4 and II the other categories 5,…,8. In the first stage, all required tables are estimated using aggregated categories for educational attainment. Then, in the second stage, tables are re-estimated using original categories for educational attainment. This can be done for the original categories 1,…,4 and 5,…,8 separately. In this way, two independent estimation problems are obtained. When estimating tables in the second step, it needs to be ensured that results are consistent with the more aggregated tables that are estimated in the first stage.

In the previous example one variable was aggregated, educational attainment. It is however also possible to aggregate multiple variables. In that case a multi-step method is obtained, in which in each stage after Stage 1, one of the variables is disaggregated.

Because of these dependencies of the estimation processes in different stages, it cannot be excluded that the three problems of Section 5.2.5 occur. However, the problems are likely to have a lower impact than in RW. This is because of a lower degree of dependency: in RW each estimated table may be dependent on all earlier estimated tables, whereas in the proposed D&C approach, estimation of a certain sub problem only depends on one previously solved problem.

5.6 APPLICATION TO DUTCH 2011 CENSUS

In this section we present results of a practical application of the proposed Divide-and-conquer (D&C) methods to the Dutch 2011 Census tables. Our aim is to test the feasibility of the methods, as well as to compare results with the officially published results that are largely based on RW. Subsection 5.6.1 describes backgrounds of the Dutch Census. Subsection 5.6.2 explains the setup of the tests and Subsection 5.6.3 discusses results.

5.6.1 Dutch 2011 Census

According to the European Census implementing regulations, Statistics Netherlands was required to compile sixty high-dimensional tables for the Dutch 2011 Census, for example, the frequency distribution of the Dutch population by age, sex, marital status, occupation, country of birth and nationality. Since the sixty tables contain many common variables, a simple linear weighting method does not lead to consistent results.

Several data sources are used for the Census, but after micro integration, two combined data sources are obtained: one based on a combination of registers and the other one is a combination of sample surveys (Schulte Nordholt *et al.*, 2011). From now on, when we refer to a Census data source, a combined data source is meant after micro integration. The 'register' data cover the full population (in 2011 over 16.6 million persons) and include all relevant Census variables except 'educational attainment' and 'occupation'. For the 'sample survey' data it is the other way around, these data cover all relevant Census variables, but it is available for a subset of 331,968 persons only.

For the 2011 Census 42 tables needed to be estimated that include ´educational attainment´ and/or 'occupation'. The target population of these tables consists of the registered Dutch population, with the exception of people younger than 15 years. Young children are excluded because the two sample survey variables 'educational attainment' and 'occupation' are not relevant for these people.

The total number of cells in the 42 tables amounts to 1,047,584, the number of cells within each table ranges from 2,688 to 69,984.

5.6.2 Setup

Below we explain how the two D&C algorithms were applied to the 2011 Dutch Census.

Setup 1 - Splitting by common variables

In this setup, the original table set is split into 48 independently estimated table sets, by using geographic area (12 categories), sex (2 categories), and employment status (2 categories) as splitting variables. Each of the 48 table sets contains a subset of the 42 Census tables, determined by the categories of the splitting variables.

The three splitting variables are however not present in all 42 Census tables. In 13 tables geographic area is missing and in one table sex is absent. Tables that do not include the three splitting variables were extended by incorporating missing variables. As a result, the total number of cells in the 42 tables was increased from 1,047,584 to 4,556,544.

Setup 2 - Aggregation and disaggregation

In this setup educational attainment (8 categories) and occupation (12 categories) were selected for aggregation of categories. Initially, both variables are aggregated into two main categories, that each contain half of the categories of the original variables. Thereafter, results were obtained for the required categories for the two aggregation variables.

Five optimization problems are defined in this procedure. In the first problem a table set is estimated based on aggregated categories for educational attainment and occupation. In each of the following stages either one of the two aggregated categories for educational attainment or occupation is disaggregated into required categories. Since less sub problems are defined, it follows that average problem size is larger than for Setup 1.

5.6.3 Results

In this subsection we compare results of the two D&C methods with the RW-based method as applied to the official 2011 Census. All practical tests were conducted on a 2.8 GHZ computer with 3.00 GB of RAM. Xpress was used as solver (Dash optimization, 2017).

A simultaneous estimation of the required 42 Census tables, as described in Section 5.4, turned out to be infeasible, due to memory problems of the computer.

The two D&C approaches were however successfully applied; there were no problems from a technical perspective and the estimation problems as experienced with RW were avoided.

Thus, we arrive at our main conclusion that the D&C approaches have broader applicability than RW.

We now continue with a comparison of the reconciliation adjustments. The criterion used to compare degree of reconciliation adjustment is based on the QP objective function in (5.7), a sum of weighted squared differences between initial and reconciled estimates, given by

$$\sum_{i=1}^{P} \frac{1}{\left(\hat{\boldsymbol{t}}_y^{w*}\right)_i} \left(\left(\hat{\boldsymbol{t}}_y^{*}\right)_i - \left(\hat{\boldsymbol{t}}_y^{w}\right)_i\right)^2 , \tag{5.9}$$

where $\hat{\boldsymbol{t}}_y^{w}$ is a vector with initial estimates, $\hat{\boldsymbol{t}}_y^{*}$ is a vector with reconciled estimates, $\hat{\boldsymbol{t}}_y^{w*} = pmax(\hat{\boldsymbol{t}}_y^{w}, 1)$. Table 5.3 compares total adjustment, as defined according to (5.9), based on all cells in all 42 estimated tables.

Table 5.3. Total adjustment; three methods

	Total adjustment	
Method	**All cells**	**Cells with initial estimate larger than zero**
Dutch 2011 Census	109.64	12.68
Splitting by common variables	88.35	12.37
Aggregation and Disaggregation	69.56	12.47

All values are $*10^6$

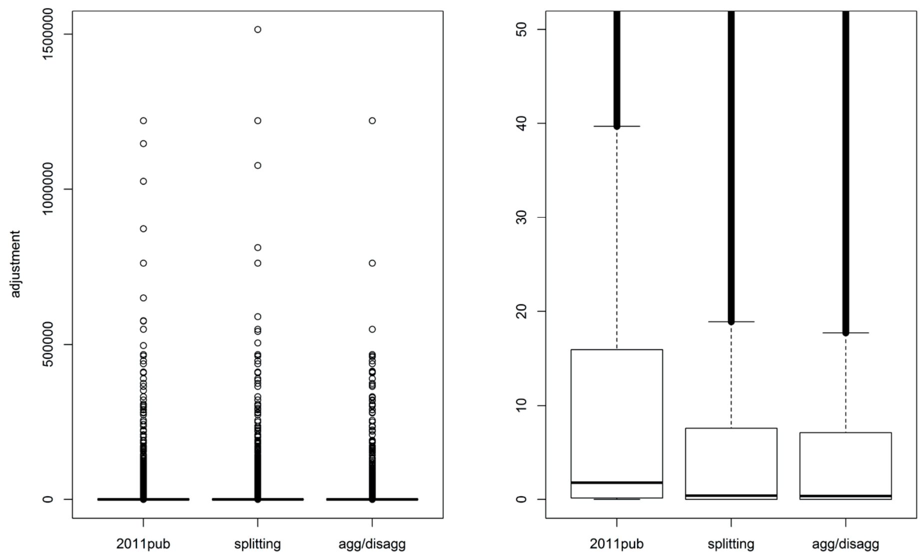

Figure 5.1 Boxplots of adjustments to all cells of 42 Census tables. The right panel zooms in on the lower part of the left panel.

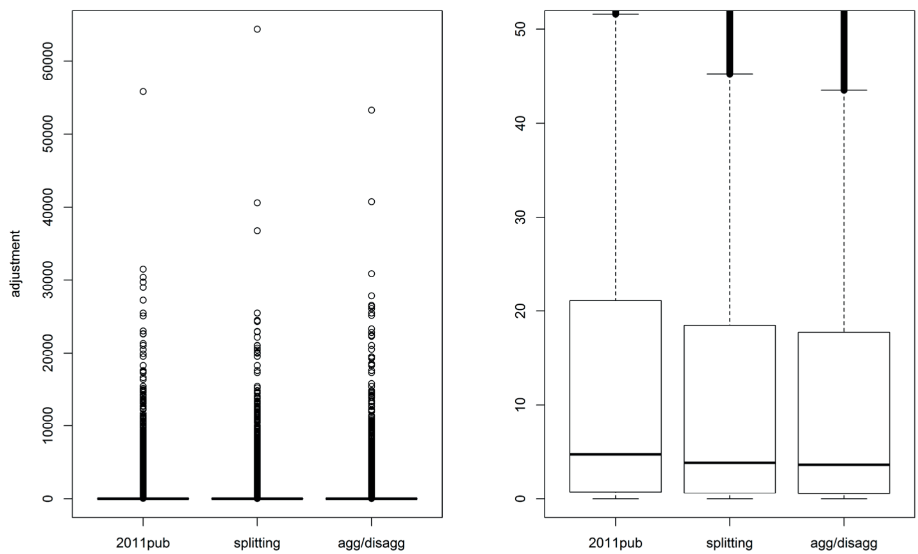

Figure 5.2 Boxplots of adjustments to all cells with nonzero initial estimates. The right panel zooms in on the lower part of the left panel.

The two newly developed D&C methods lead to smaller total adjustment than RW. The result that "Aggregation and Disaggregation" method gives rise to a better solution than "Splitting by common variables" can be explained by the lower amount of sub-problems that are defined in the chosen setups. If we only compare cells with larger than zero initial estimates, differences between three methods become very small. This shows that the way how original estimates of zero are processed is more important than the way how the estimation problem is broken down into sub problems.

The boxplots in Figures 5.1 and 5.2 compare adjustment at the level of individual cells. It can be seen that the amount of relatively small corrections is larger for the two D&C methods than for the RW-based method used for officially published Census tables. Differences in results are however smaller again, if zero initial estimates are not taken into account.

5.7 DISCUSSION

When several frequency tables need to be produced from multiple data sources, the problem may arise that results are numerically inconsistent. That is, that different results are obtained for the same marginal totals in different tables. To solve this problem, Statistics Netherlands developed a Repeated Weighting (RW) method. This method was applied to the 2001 and 2011 Dutch censuses. However, the scope of applicability of this method is limited by several known estimation problems. In particular, the sequential way of estimation causes problems. As a result of these problems, estimation of the 2011 Census was troublesome. A suitable order of estimation was found after long trial and error.

This chapter presented two alternative Divide and Conquer (D&C) methods that break down the estimation problem as much as possible into independently estimated parts, rather than the dependent parts that are distinguished in RW. One of the two newly developed methods partitions a given table set according to common categories of variables that are contained in each table. The other method is based on aggregation and disaggregation of categories. As a result of independent estimation, many of the estimation problems of RW can be prevented. This greatly enhances the practicability of the method. Another advantage is that the reduced order-dependency of results leads to less ambiguity. A final advantage is that the new approaches can be more easily extended to incorporate additional requirements, like non-negativity of estimates and solutions for the empty cell problem.

An application to 2011 Census tables showed that estimation problems were actually avoided. Reconciliation adjustments were observed to be smaller than in RW. Hence, it can be argued that a better solution can be obtained that deviates less from the data sources. The smaller adjustments can be mainly attributed to the solution applied to the empty cell problem; a solution that could not be implemented in the original RW approach. Hence, the key message of this chapter is when estimating a consistent set of tables, there can be smarter ways of breaking down the problem than estimating single tables in sequence.

For problems in which a simultaneous estimation of all tables is computationally feasible, such an approach is to be preferred. Most importantly, because a full simultaneous approach avoids the estimation problems that are experienced with RW. Moreover, an optimal solution is obtained with minimal adjustment from the data sources. Finally, from a practical point of view solving one (or few) optimization problem(s) is much easier than solving many problems.

6. Constraint simplification for data editing of numerical variables[6]

Summary. Data editing is the process of checking and correcting data. In practice, these processes are often automated. A large number of constraints needs to be handled in many applications. This chapter shows that data editing can benefit from automated constraint simplification techniques. Performance can be improved, which broadens the scope of applicability of automatic data editing. Flaws in edit rule formulation may be detected, which improves the quality of automatic edited data.

6 This chapter has been published as: Daalmans J.A. (2018) Constraint Simplification for Data Editing of Numerical Variables, *Journal of Official Statistics*, 34, 27-39.

6.1 INTRODUCTION

Collected micro data usually contain errors, e.g. pregnant men, average salary of 5 million euro and components of a total that do not add up to that total. Correction of such errors is often necessary to prevent flaws and inconsistencies in statistics to be published. The process of checking and correcting is called data editing, see e.g. De Waal *et al.* (2011) and Pannekoek *et al.* (2013). A common approach for data editing is based on the paradigm of Fellegi and Holt (1976), stating that the data in a record should be adapted to satisfy all edit rules by changing the fewest possible values.

Error localization according to the Fellegi and Holt paradigm can be formulated as a Mixed Integer Linear Program (MILP) problem, see e.g. De Waal *et al.* (2011). Although, in general, a solution to a data editing problem can be found in reasonable time - typically a few seconds - the worst-case performance of a MILP problem is known to be exponential in the problem size. Even when using modern computers, it can take hours or even days to obtain a solution for a single record. From Operations Research and Artificial Intelligence it is well known that performance of a mathematical optimization problem can be improved by a constraint simplification step, see e.g. Paulraj and Sumathi (2010), Telgen (1983), Chmeiss *et al.* (2008) and Piette (2008). This means eliminating redundant constraints and simplifying unnecessary complicated constraints, before optimization. Nevertheless, remarkably few applications of constraint simplification are known in the context of data editing. Bruni (2005) explains how redundant edit rules can be detected. Also, Statistics Canada developed a software tool with simplification features (BANFF support team 2008, Chap. 2). These applications do however not allow for conditional ("IF-THEN") rules, where the variables involved in the IF and THEN statements may contain errors. Such rules frequently occur in official statistics and are especially problematic for computational performance, due to the integer variables that arise in the corresponding MILP problem.

This chapter contributes to fill the gap for constraint simplification techniques for error localisation of numerical data. Special attention will be given to conditional rules. We present automated methods that work at the formal level through solving MILP problems. An advantage of automated procedures is that removal of redundant constraints can be done out of sight, so that users can still specify all possible rules without ending up with an inefficient edit set. Working at the formal level means that the methods can be applied to a generic class of rules, regardless their semantic meaning.

Since edit rule simplification improves computational performance, it has the potential of further extending the application possibilities of automated data editing. Besides this, expert's feedback on automatic detected redundant edit rules might help to increase the understanding of the joint effects of a set of rules. Due to the complex interdependence and misspecification, a set of rules may have different implications than intended. Correction of erroneous rules improves the quality of automatic edited data and avoids the need for manual adjustment of results. For example, the following redundant rule was found in an edit set, actually used by Statistics Netherlands:

$$\begin{aligned}&\text{IF (Questionaire\_ID} \neq 1 \text{ OR Questionaire\_ID} \neq 2\text{) THEN}\\&\text{(VariableX = VariableY)}\end{aligned} \tag{6.1}$$

Manual inspection might reveal that the OR-operator was meant to be an AND-operator.

The structure of this chapter is as follows. Section 6.2 describes the MILP formulation for data editing of numerical variables. Sections 6.3-6.5 present formal, mathematical algorithms for simplifying edit sets: eliminating single variables is described in Section 6.3, eliminating redundant parts from conditional rules is discussed in Section 6.4 and the redundancy of rules as a whole is considered in Section 6.5. Section 6.6 presents real-life applications of constraint simplification and data editing. Finally, Section 6.7 finishes this chapter with a discussion.

6.2 OUTLINE OF BASIC APPROACH

We introduce the basic idea of MILP problems first. Then, it is explained how edit rules can be translated into MILP constraints.

6.2.1 Definition of a MILP problem

A MILP problem consists of a loss function to be minimized and a set of inequality constraints involving both real and integer variables. A general form is given by

$$\begin{aligned}&\text{Minimize } f(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{c}^{\mathrm{T}}\binom{\boldsymbol{x}}{\boldsymbol{z}},\\&\text{s.t. } \mathbf{A}\binom{\boldsymbol{x}}{\boldsymbol{z}} \leq \boldsymbol{b},\\&\boldsymbol{x} \in \mathbb{R}^{p} \text{ and } \boldsymbol{z} \in \mathbb{Z}^{q}\end{aligned} \tag{6.2}$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ are vectors of real and integer decision variables, $\mathbf{c}$ is a constant vector ($\boldsymbol{c} \in \mathbb{R}^{p+q}$), $\mathbf{A}$ is a coefficient matrix and $\boldsymbol{b}$ a vector of upper bounds, see e.g. Bertsimas and Tsitsiklis (1997).

In the remainder of this chapter several algorithms are proposed that make use of the feasibility of a set constraints. This can be checked by a MILP solver by using a trivial loss function with $\boldsymbol{c} = 0$. Of course, if a solution exists the optimum value will be zero, but if the set of constraints is infeasible, most MILP solvers return an error message.

6.2.2 Edit rules as MILP constraints

This subsection introduces the edit rules that are considered in this chapter and explains how these rules can be transformed into MILP constraints. The edit rules in this chapter can be subdivided into unconditional and conditional rules.

We consider linear unconditional rules, like

Total turnover = Domestic turnover + Foreign Turnover,
Total turnover ≥ 0,

that can be straightforwardly formulated as MILP contraints. One could note that the constraints in (6.2) do not allow for "larger than" and "equality" signs, but it is well-known that these rules can be reformulated into the required form. For example, an equality can be written as two inequalities and a constraint $x > 0$ can be approximated by $-x \leq -\varepsilon$, where ε is a sufficiently small value.

We also consider 'simple' and ' compound' conditional rules. A 'simple' conditional edit has the following form

IF <Statement 1 > THEN <Statement 2>,

where each "statement" is a linear equality or inequality. Compound rules may also contain:

- AND-operators in the IF-clause and/or
- OR-operators in the THEN-clause.

An example is:

IF (Number of employees > 0 AND Turnover > 0) THEN (Wages > 0
OR Labour costs > 0). (6.3)

Note that above we did not consider rules with:

- OR-operators in the IF-clause and/or
- AND-operators in the THEN-clause,

but these rules can be rewritten as a number of simple conditional rules. For example, the edit:

IF (Number of employees > 0 OR Turnover > 0) THEN
(Wages > 0 AND Labour costs > 0)

is equivalent to the combination of the following four "simple" conditional rules:

IF Number of employees > 0 THEN Wages > 0,
IF Number of employees > 0 THEN Labour costs > 0,
IF Turnover > 0 THEN Wages > 0,
IF Turnover > 0 THEN Labour costs > 0.

As mentioned by Chen *et al.* (2010), conditional rules need to be expressed in Disjunctive Normal Form (DNF), before these can be further converted into the required MILP format. A DNF is a disjunction of assignments (a sequence of OR's) that makes a rule True, see e.g. Hooker (2000).

To explain the transformation to DNF, note that a conditional rule is satisfied, if either the IF-clause is violated, or if the THEN-clause is fulfilled. Thus, a condition rule can be put in DNF, by joining the negation (i.e. opposite) of the "IF"-clause with the original "THEN"-clause. For example, the rule: "If Turnover > 0 THEN Wages > 0" can be stated as "Turnover ≤ 0 OR Wages > 0".

For compound edits, the IF-clause is assumed to be a conjunction (sequence of AND's). According to Morgan's law, the negation of a conjunction is a disjunction of negations. To illustrate this, the example in (6.3) can be written in DNF as

Number of employees ≤ 0 OR Turnover ≤ 0
OR Wages > 0 OR Labour costs > 0,

where the first two statements are negations of the original IF-clause statements.

An expression for n_C edit rules in DNF is given by

$$\bigcup_{j=1}^{D_i} \left(\left(a_{ij}^C\right)^T \boldsymbol{x} \leq b_{ij}^C \right) \qquad i = 1, \ldots, n_C. \tag{6.4}$$

where an edit rule i is stated as a disjunction with D_i disjunctive terms. The coefficients and upper bounds for the jth term are denoted by a_{ij}^C and b_{ij}^C respectively. Again, 'equality', 'larger than' or 'smaller than' constraints can be reformulated into the form (6.4).

To express the constraints in (6.4) as MILP constraints, the following formulation can be used, based on the so-called Big M method.

$$\begin{aligned} \left(a_{ij}^C\right)^T \boldsymbol{x} &\leq b_{ij}^C + M(1 - z_{ij}), && i = 1, \ldots, n_C, \quad j = 1, \ldots, D_i, \\ \sum_{j=1}^{D_i} z_{ij} &= 1 && i = 1, \ldots, n_C, \\ -z_{ij} &\leq 0 && i = 1, \ldots, n_C, \quad j = 1, \ldots, D_i. \end{aligned} \tag{6.5}$$

where z_{ij} are integer variables and M is a sufficiently large constant.

The equation $\sum_{j=1}^{D_i} z_{ij} = 1$ guarantees that only one disjunctive term is selected per disjunction. For each selected term (i, j with $z_{ij} = 1$), it is enforced that $\left(a_{ij}^{C}\right)^{T} \boldsymbol{x} \leq b_{ij}^{C}$. For each non-selected term (i, j with $z_{ij} = 0$), the first constraint in (6.5) becomes redundant.

As shown in (6.5) integer variables are needed for the formulation of conditional rules. Because integer variables are much less efficiently handled than continuous variables, conditional rules can be less efficiently processed than unconditional ones. Therefore it is very beneficial to replace conditional rules by unconditional ones.

6.3 FIXED VALUE ELIMINATION

The aim of this technique is to shorten edit rules by elimination of 'fixed' variables, i.e. variables with only one admissible value. As a result, an edit set may become simpler, possibly giving rise to a better performance of data editing software. Moreover, misspecification of edit rules might be detected by manual inspection of fixed values. Consider the following example:

$$\begin{aligned} &\text{Edit 1: } x_1 + x_2 + x_3 = 10, \\ &\text{Edit 2: } x_1 + x_2 \geq 10, \\ &\text{Edit 3: } x_3 \geq 0. \end{aligned} \tag{6.6}$$

It is immediately clear that x_3 necessarily has to be zero. In other words, x_3 is a fixed variable.

Fixed values can be identified by solving two MILP programming problems for each continuous variable. Each variable is minimized and maximized once, subject to the MILP representation of the edit rules. If the minimum and maximum value turn out to be the same, the variable at hand is a fixed variable. Its value can be substituted in all edits in which it appears and a constraint is added stating that the fixed variable can only attain the fixed value.

Besides fixed values, any finite minimum or maximum is a candidate for content-wise analysis, because these bounds may be different than intended.

In our example, we can add the rule $x_3 = 0$ to our edit set and substitute x_3 in all other rules. We obtain

$$\begin{aligned} &\text{Edit 1'}: x_1 + x_2 = 10, \\ &\text{Edit 2'}: x_1 + x_2 \geq 10, \\ &\text{Edit 3'}: 0 \geq 0. \\ &\text{Edit 4'}: x_3 = 0. \end{aligned} \tag{6.7}$$

6

Of course, these edits can be further simplified, Edits 2' and 3' are obviously redundant. The further simplification for redundant rules will be explained in Section 6.5.

6.4 SIMPLIFICATION OF COMPOUND RULES

This section deals with the simplification of compound rules by elimination of unnecessary statements. Two new MILP algorithms are presented, based on existing methods from Dillig *et al.* (2010). The aims are again to improve computational performance and to detect misspecification of edit rules. A possible outcome, especially beneficial to computation performance, is that a conditional rule can be replaced with an unconditional one.

6.4.1 Implicitly unsatisfiable statements

In this subsection compound edit statements of the form (*A* OR *B* OR …) are simplified by deletion of statements that cannot be satisfied, given the available set of edit rules. Dillig et *al.* (2010) call these statements "non-relaxing", since these do not enhance the feasible area of a MILP problem. If, after simplification, only one component remains, a conditional rule has been converted into an unconditional one. An example is

Edit 1: $x_1 > 0$ OR $x_2 > 0$,
Edit 2: $x_2 < 0$.

It is immediately clear that the statement $x_2 > 0$, within Edit 1, cannot possibly be satisfied, because of Edit 2. This statement can be removed from Edit 1, since it is redundant. Consequently, Edit 1 can be formulated as an unconditional rule. A more formal definition is given below:

Definition

A statement e_{ij} of a compound edit e_i within a feasible edit set $\mathbf{E}$ is implicitly unsatisfiable, if $\mathbf{E} \cup e_{ij}$ is infeasible.

Here, $\mathbf{E} \cup e_{ij}$ stands for the edit set that is obtained by extracting a compound edit's component e_{ij} from e_i and adding it to the set $\mathbf{E}$, as if it were a single edit. An algorithm for removal of implicitly unsatisfiable statements is stated below

Algorithm 1: Identification & removal of implicitly unsatisfiable statements

Input: Feasible edit set **E**
Output: Feasible edit set **E**, without implicitly unsatisfiable components.
For each compound edit $e_i \in \mathbf{E}$ do
  For each statement $e_{ij} \in e_i$ do
    $\mathbf{E}^* \leftarrow \mathbf{E} \cup e_{ij}$;
    IF isFeasible($\mathbf{E}^*$) = FALSE THEN $e_i \leftarrow e_i \setminus e_{ij}$
  Next
Next

In each step one statement of a compound edit is added to a feasible edit set. Subsequently, the feasibility of the extended edit set is checked by isFeasible(), a function that can be implemented by a MILP solver, see Section 6.2. If the extended edit set is infeasible, the compound edit's statement is implicitly unsatisfiable and therefore redundant.

When applied to our previous example, the algorithm means that the constraints $x_1 > 0$ and $x_2 > 0$ are added to Edits 1 and 2 one by one and that the feasibility is verified for both resulting edit sets. Because the addition of $x_2 > 0$ renders Edits 1 and 2 infeasible, $x_2 > 0$ is an implicitly unsatisfiable statement. It can be deleted from Edit 1 accordingly.

6.4.2 Implicitly satisfied statements

This subsection aims at replacing compound edit rules (*A* or *B* or…) with single, unconditional rules. The main idea is that if a compound rule contains a statement (say *A*) that is necessarily *True*, the compound rule can be replaced with that single statement. Implicitly satisfied statements are called non-constraining by Dillig *et al.* (2010), since these do not reduce the feasible region of a MILP problem. Consider the following example:

Edit 1: $x_1 < 50$ OR $x_2 > 100$,
Edit 2: $x_1 > 100$ OR $x_2 > 0$.

For all possible x_1 values, at least one of the statements $x_1 < 50$ and $x_1 > 100$ is not satisfied. Thus, Edits 1 and 2 imply that either $x_2 > 0$, or the even stronger condition $x_2 > 100$, needs to be true. As a result, we obtain that $x_2 > 0$ always needs to hold, in other words $x_2 > 0$ is implicitly satisfied. Consequently, Edit 2 can be replaced with this single statement. A more formal definition is stated below:

Definition

A component e_{ij} of a compound edit e_i within a feasible edit set **E** is implicitly satisfied if $\mathbf{E} \cup \neg e_{ij}$ is infeasible (where ¬ stands for negation).

This definition makes use of the equivalence of the statements that a compound edit's component is implicitly satisfied and that the opposite of that component cannot occur. An algorithm for identifying implicitly satisfied statements is as follows

Algorithm 2: Identification & replacement of implicitly satisfied statements

Input: Feasible edit set **E**
Output: Feasible edit set **E**, without implicitly satisfied statements.
FOR each compound edit $e_i \in \mathbf{E}$ DO
    FOR each statement $e_{ij} \in e_i$ DO
        $\mathbf{E}^* \leftarrow \mathbf{E} \cup \neg e_{ij}$;
        IF isFeasible($\mathbf{E}^*$) = FALSE THEN $\mathbf{E} \leftarrow \{\mathbf{E} \backslash e_i\} \cup e_{ij}$
    NEXT
NEXT

This algorithm has a similar structure as Algorithm 1. Each step of the algorithm checks the feasibility of an extended edit set that is obtained by adding the negation of a statement of a compound rule to the given edits in **E**. If the resulting edit set turns out to be infeasible, the added statement is "implicitly satisfied". The statement is added to the edit set as an unconditional rule and the conditional rule from which the statement is obtained is deleted.

When applied to our previous example, the constraints $x_1 \geq 50$, $x_2 \leq 100$, $x_1 \leq 100$ and $x_2 \leq 0$ are added to Edits 1 and 2 one by one, which are the negations of the original edit components. Feasibility is checked for all resulting edit sets. Because the addition of $x_2 \leq 0$ renders Edits 1 and 2 infeasible, $x_2 > 0$ is implicitly satisfied. Hence, Edit 2 can be replaced with the unconditional rule $x_2 > 0$.

6.5 REDUNDANT EDIT REMOVAL

This subsection's aim is to simplify edit sets by removal of redundant edits, i.e. rules that can be left out of an edit set, without affecting the set of feasible records. The removal of redundant constraints may speed up the error correction process without losing power of correction. Because redundant edits may emerge as a result of fixed value substitution and simplification of conditional edits, it is important that redundant edit removal is conducted after these other steps. Consider the following example:

Edit 1: $x_1 + x_2 \leq T_1$,
Edit 2: $x_3 + x_4 \leq T_2$,
Edit 3: $T_1 + T_2 = T_3$,
Edit 4: $x_1 + x_2 + x_3 + x_4 \leq T_3$.

Edit 4 can be omitted because it is implied by Edits 1, 2 and 3.

An edit is redundant if other edits imply that the edit is 'always satisfied'. As mentioned in Subsection 6.4.2, this is equivalent to the statement that the negation of the edit cannot occur. This leads to the following definition

Definition

An Edit e_i from an edit set **E** is redundant, if $\{\mathbf{E} \setminus e_i\} \cup \neg e_i$ is infeasible.

The edit set $\{\mathbf{E} \setminus e_i\} \cup \neg e_i$ is obtained from **E**, by replacing Edit e_i by its negation.

In literature many methods have been mentioned for detection of redundant constraints. Paulraj and Sumathi (2010) performed a comparative study. Below we describe a method mentioned by e.g. Felfernig *et al.* (2011), Chmeiss *et al.* (2008) and Bruni (2005). The reason for choosing this method is its simplicity and the possibility of implementing it by a MILP solver.

Algorithm 3: Identification & removal of redundant edits

Input: Feasible edit set **E**

Output: Feasible edit set **E**, without redundant edits

1 FOR each Edit $e_i \in \mathbf{E}$ DO

2 $\mathbf{E}^* \leftarrow \{\mathbf{E} \setminus e_i\} \cup \neg e_i$;

3 IF isFeasible($\mathbf{E}^*$) = FALSE THEN $\mathbf{E} \leftarrow \mathbf{E} \setminus e_i$

4 NEXT

When applied to previous example, the algorithm means that the negations of Edits 1, 2, 3 and 4 are added to the edit set one by one and that the feasibility is verified for all of the resulting set of rules. In this way, the redundancy of Edit 4 can be easily demonstrated.

Below a few words on the computation of negations. The negation of an equality constraints can be expressed as combination of two inequality constraints. For example, in previous example the negation of Edit 3, can be expressed as $T_1 + T_2 < T_3$ OR $T_1 + T_2 > T_3$. These two constraints are added to the three original rules one by one. Only if both additions lead to infeasible edit sets, one could conclude that Edit 3 is redundant. In our example, Edit 3 is however clearly not redundant.

The negation of a compound edit rule e_i, expressed as the disjunction

$$\bigcup_{j=1}^{D_i} \left(\left(a_{ij}^C\right)^T \boldsymbol{x} \leq b_{ij}^C \right),$$

is given by,

$$\left(\left(a_{ij}^C\right)^T \boldsymbol{x} > b_{ij}^C \right) \qquad\qquad j = 1, \ldots, D_i,$$

a combination of D_i linear, unconditional constraints that all have to be satisfied.

6.6 APPLICATIONS

Aim of this section is to apply constraints simplification methods on 'real-life' edit sets. We would like to show that these edit sets can actually be simplified. Moreover, we demonstrate that constraint simplification improves data editing's performance.

All applications were performed on a 32-bit Windows 7 desktop with a 2.80 GHz CPU and 3 Gigabyte of RAM memory. The methods from Chapters 3-5, were implemented by R. The free LpSolveAPI was used as a MILP solver (Konis, 2016) and the Editrules package (De Jonge and Van der Loo, 2015) was implemented for automatic data editing. The following edit sets were considered

1. Sales: Real-life edit set used for the 2012 Dutch Structural Business Statistics for sale of motor vehicles for businesses with fewer than 10 employed persons;
2. Maintenance: Real-life edit set used for the 2012 Dutch Structural Business Statistics for maintenance of motor vehicles for businesses with fewer than 10 employed persons;
3. Health-care: Edit set under development, meant to be used for a Dutch survey among welfare and childcare institutions;

All methods for constraint simplification in Sections 6.3-6.5 were applied to these three data sets. Automatic data editing was applied to the first two edit sets only, because of the lack of data for the third application.

Table 6.1. Results of three real-life applications

	Sales	Maintenance	Health-care
Original edits			
Number of edits	115	119	196
-of which conditional:	26	29	114
Number of variables in edits	74	74	75
Simplification			
Fixed values	3	7	2
Conditional edits			
-Implicitly unsatisfiable components	1	1	4
-Implicitly satisfied components	1	1	3
Redundant edits	22	29	10
-of which conditional	7	13	3
Cleaned edits			
Number of edits	93	90	186
-of which conditional:	19	16	104
Computation Time (in seconds)	5	6	2,465

Firstly, Table 6.1 shows the feasibility of constraint simplification on a regular computer. One could note that computation time for the third application is relatively large, about 40 minutes, which can be explained from the many conditional rules. Large computation time is however not a problem, because edit rules simplification only needs to be conducted once, after designing or revising an edit set.

Secondly, Table 6.1 demonstrates that all simplification features in Sections 6.3-6.5 are useful, as each feature actually simplifies all three edit sets. The total number of edit rules is reduced by 5-20%; the number of conditional edits by 10-45%.

Table 6.2 shows that total computation time for automatic data editing is reduced by 23% for the Sales application and even by 55% for the Maintenance data set. The latter reduction can be largely attributed to only one record, whose computation times are 297 and 109 seconds for the original and simplified edit sets. This actually points out that the worst-case performance is important in data editing, but also shows that worst-case performance can be noticeably improved by rule simplification.

A practical solution to the possibly long computation time is to limit the available time for each record. The last row in Table 6.2 shows that edit rule simplification slightly increases the amount of records that are processed within 10 seconds.

Table 6.2 Automatic data editing, original and simplified edit sets

	Sales (N=614 records)		**Maintenance (N=197 records)**	
	Original edits	**Simplified edits**	**Original edits**	**Simplified edits**
Processed records*	613	613	197	197
Total time (in seconds)	2,639	2,039	479	217
Records processed within 10 sec.	592	598	191	194

*= given a maximum computation time of 10 seconds per record.

6.7 DISCUSSION

Many works from the literature present automatic constraint simplification techniques that are able to greatly improve computational performances of large optimization problems. But, to the best knowledge of the author, these techniques are not often applied in the field of data editing.

This chapter shows that automated data editing can benefit from constraint simplification. A number of methods was presented for numerical data, based on MILP programming. Much attention was given to conditional IF-THEN rules that often occur in official statistics and that are particularly important for computational performance.

The feasibility of constraint simplification was demonstrated on a regular computer using freely available MILP solvers. It was shown that real-life edit sets can actually be simplified. As a result, the total computation time for localising erroneous values was reduced up to 55%; a reduction that can be mainly attributed to a few records with the largest computation time. Hence, constraint simplification is an important step in further enhancing the practicality of automatic data editing.

Another benefit is that constraint simplification provides insight in the joint consequences of a set of rules. Manual inspection of automatically determined redundant rules and variables with a fixed value or finite bounds might reveal errors in rule formulation. Correction of these errors increases the quality of automated data editing and reduces the need for manual correction of automatically edited data.

A practical merit of the proposed methods is that simplification can be automated, out of sight of users, so that practitioners in the field do not have to bother about specifying constraints in a compact way.

This chapter implicitly assumed that edits are interconnected. However, if this is not the case, it is advisable to split an edit set $\mathbf{E}$ into disjunct sets, $\bigoplus_i \mathbf{E}_i$, such that $e_i \in \mathbf{E}_i$ and $e_j \in \mathbf{E}_j$ $(i \neq j)$ do not have any variable in common. Disjunct edit sets can be treated independently, which may improve performance of both data editing and edit rule simplification.

The simplification methods in this chapter have been designed for feasible edit sets. Despite that infeasible edit rules are useless for practical application, infeasible rules may occur in practice, for instance due to misspecification. In general, it can be hard to find the cause of a contradiction, especially if the number of edit rules is large. Therefore, most methods for dealing with inconsistency concentrate on isolating a smallest possible subset of inconsistent edit rules: a so-called irreducible inconsistent subset (IIS). Several algorithms for detecting IIS's are available from literature. The so-called "Deletion Filter" by Chinneck (1997) can be advised for many applications as it is easily understood, suitable for conditional "IF-THEN" edits and applicable for MILP programming. In a recent publication, Bruni and Bianchi (2012) proposed another, innovative approach, based on Farka's lemma. Their method however relies on an assumption, the so-called Integral Point property, that is unknown to be true for general applications.

A direction for further research is to introduce more constraint simplification techniques for data editing. In this chapter we considered numerical data. Methods for categorical data could be developed in the future..

7. Discussion

This thesis' aim is to push the boundaries of the application possibilities of formal macro integration methods. Since their development in the 1940s formal macro integration methods have been increasingly used (Schneider and Zenios, 1990). Traditionally, these methods have been applied to national accounts reconciliation. Compared to traditional informal approaches, formal macro integration methods increase transparency, reduce subjectivity and save resources. Currently, macro integration methods are often used in the development of new statistical output from multiple data sources. In a recent study, Cuevas *et al.* (2015) presented a macro integration approach to compile quarterly gross domestic product (GDP) at the regional level. Tukker and Dietzenbacher (2013) use macro integration in the construction of global multi-regional input–output data bases (GMRIO): very detailed supply and use tables of many countries, supplemented with data on air emissions and resource use, see e.g. Stadler *et al.* (2018) and Canning and Wang (2005). A growing tendency can be observed to use macro integration methods outside the field of National Accounts. For instance, De Beer *et al.* (2010) report on an application to social statistics: the problem to achieve harmonised estimates of migration flows, where sending countries report different numbers than receiving countries. Uses of macro integration methods are also found in fields that are completely different from official statistics. In chemical engineering, for example, the problem occurs that several measurements are made that have to satisfy certain theoretical relations, but that initially fail to do so. A vast amount of papers have been published that propose methods to achieve consistency, amongst others Tamhane and Mar (1985) and Özyurt and Pike (2004). The ongoing development of new methods, the drastically increased automation possibilities and the rapid access to new data sources offers increasingly more possibilities for setting up consistent statistics that could not be compiled before.

7

The chapters of the current thesis are inspired by implementation problems that were actually observed at Statistics Netherlands. The main contribution is threefold: 1) To develop new methods; 2) to further explore the properties of existing methods and 3) to facilitate the disclosure to new application areas. Section 7.1 summarizes the main contributions, Section 7.2 exposes ideas for further research.

7.1 CONTRIBUTION OF THE THESIS

The subsection presents an overview of the main contributions per chapter.

Chapter 2 extends an existing multivariate Denton method to include a broad class of modelling possibilities that have already been available for Stone's methodology. The features include: linear and ratio constraints, hard and soft constraints, weights and different objectives for different series. The new model is based on a flexible quadratic programming approach that can be easily tailored to the user's needs. The method has been implemented in the production process of Dutch National Accounts, but can be more generally applied for any multivariate benchmarking problem, in which relations between time series are of key interest.

Chapter 3 enriches the current insights on the usability of the Growth Rate Preservation (GRP) method for benchmarking; a method that is preferred by some works in the literature. It shows that GRP does not satisfy a 'time reversibility' property. This property is well-known from index theory, but has not been mentioned before in the context of benchmarking. Time reversibility means that the results of any benchmarking method should be independent of time direction. Methods that do not satisfy this property give rise to an undesirable arbitrariness and might even change the timing of the most important economic events. Therefore, any benchmarking method should preferably satisfy time reversibility. Because of this and because of other undesirable properties, Denton can be considered a better alternative than GRP for many benchmarking applications.

Chapter 4 provides a solution to a "sequential estimation" problem; a relevant practical problem that is often neglected in the literature. Sequential estimation is needed if benchmarking cannot be applied to all available past data, but only to the most recent part of a time series. The motivation for this restriction is purely practical: data for the distant past are often not subject to change, as these have already been published. It has been demonstrated that sub optimal movement preservation is a problem for a sequential application of Denton. A new solution has been proposed that improves on a Denton method.

Chapter 5 responds to the growing tendency to apply macro integration techniques outside the field of National Accounts. Stone's macro integration method is applied for the compilation of the Dutch Population census. It has been shown that some estimation problems, as experienced by the previously used weighting methods, can be avoided in the macro integration approach. Due to the large size of the estimation problem, it is proposed to use a so-called Divide-and-conquer approach. This means that the estimation problem is split into smaller sub parts. A similar approach can also be useful for other large data integration problems.

Chapter 6 deals with a problem that becomes relevant after implementation of any data correction method: the problem to setup and maintain a set of constraints. Several known methods from Artificial Intelligence and Operations Research have been proposed to simplify a set of constraints. It has been demonstrated that constraint simplification improves the performance of data reconciliation software and that these techniques can be used to detect errors in the formulation of constraints. To facilitate their use, the methods from Chapter 6 have been implemented in the R-package "Validate tools" (De Jonge *et al.*, 2018).

7.2 FURTHER RESEARCH

This subsection presents five directions for the further development of macro integration methods.

A first topic for future research is the further extension of macro integration methods with new functionalities. The benchmarking model in Chapter 2 extends an existing method, but it can be further extended in the future. For example, conditional ("IF-THEN") rules that are discussed in Chapter 6 for the data editing problem can be incorporated in the benchmarking model of Chapter 2. Oppositely, features that are described in Chapter 2 can be useful for data editing. For instance, Scholtus (2013) proposed to use soft constraints for data editing. Due to the close similarities between data editing and macro integration methods, both problems can profit from an interchange of ideas.

A second topic for further research is to further elaborate the desired properties of macro integration methods. The model that is proposed in Chapter 2 meets existing and newly described properties. Chapter 3 also presents properties that are essential for any benchmarking method. The need for certain properties might become clear after the implementation of a method. After the introduction of the benchmarking method from Section 2 in Statistics Netherlands' production processes, the definition of the weight expression, as presented in Section 2, has been changed to better serve users' needs. In Section 2 all expressions for the weights have been chosen proportional to the squared values of the variables. After further consideration, it turned out to be better to choose weights that are proportional to the absolute values. This change made the model more invariant under aggregation. This property means that aggregating benchmarked data or benchmarking aggregated data leads to the same results. Or, in other words, that the level of aggregation is unimportant. The reader is referred to Brent (2016) for more details.

7

A third topic for further research is the determination of weights. The weights in a macro integration model are an important means to influence results. The relative values for the weights determine which variables are adjusted the most. From a theoretical point of view, weights should capture all sampling and non-sampling errors (Stone, 1942). In practice, measurement of these errors turns out difficult. Non-sampling errors are especially problematic. In a recent study, Yingfu *et al.* (2017) put efforts to measure all errors for an application to Swedish national accounts data. Their approach still relies on many implicit, unverifiable assumptions. Like many other macro integration models, the model in Chapter 2 uses 'subjective' weights; weights that rely on expert assessments of reliability. A practical advantage of such an approach is its flexibility. A drawback is the subjectivity in it. Alternative 'objective' methods for determining weights are available from the literature. For instance, variances can be estimated from historic reconciliation corrections (Weale, 1985) or from the variability of the data in previous periods (Weale, 1992). A comparative study of different methods to generate weights can be an interesting topic for further research.

A fourth topic for further research relates to bias detection. Formal macro integration methods assume that the data are free of error. Therefore, it is required that data are cleaned from bias prior to data reconciliation. Bias detection is usually done by confronting the preliminary data with constraints. Alternatively, reconciled data can be compared with preliminary data. Formal procedure for bias detection have been implemented in the production process of Dutch National Accounts, based on ranking the discrepancies and corrections. Methods for bias detection can also be borrowed from chemical engineering. Several statistical tests have been mentioned in Tamhane and Mar (1995) and the references therein. A feasibility study of these methods for national account reconciliation can be a subject for further research. As an alternative to removing bias prior to data integration, one might consider to correct for errors and the remaining inconsistencies in one step. The chemical engineering literature mentions several methods that are designed for that purpose. Such methods are often based on the ideas of robust regression: implausible estimates are automatically attached a lower weight in the optimization, so that their influence on other variables is reduced. The meaningfulness for national account reconciliation is doubtful. On the one hand, it is desirable that one error does not affect results of other variables. On the other hand, the plausibility of the results for the variables with erroneous measurements is unclear. Whenever plausible results can be derived from the other measurements in the system, the method might work. Otherwise, the use of expert knowledge for error correction seems inevitable. Further research might provide more insight into the usefulness of these methods.

A fifth and last idea for further research is to further develop methods for sequential estimation. Sequential estimation of relatively short time series is often observed in practice, but not much attention is given for this problem in the literature. Chapter 4 presents a solution for a sequential benchmarking method. The proposed solution can be further elaborated in the future, for instance by making use of sophisticated prediction methods from time series analysis. Furthermore, a Denton method is used as a starting point in Section 4. However, for sequential estimation, a simple method like pro rata might be more appropriate.

8. References

B

Bakker, B.F.M., J. van Rooijen, L. van Toor (2014) The system of social statistical datasets of Statistics Netherlands: an integral approach to the production of register-based social statistics. *Statistical journal of the IAOS*, 30, 411-424.

Banff Support Team (2008) *Functional Description of the BANFF System for Edit and Imputation.* Technical report, Statistics Canada, Ottawa.

Battista, V., C. Manning and K.W. Robertson (2009) Research on Quarterly Benchmarking for the Current Employment Statistics Survey. In: Proceedings of Government Statistics Section, American Statistical Association, Washington, 2866–2876.

Beaulieu, J.J. and E.J. Bartelsman (2006) Integrating Expenditure and Income Data: What to Do with the Statistical Discrepancy? In: Jorgenson D.W., J.S. Landefeld and W. D. Nordhaus (eds.) A New Architecture for the U.S. National Accounts, University of Chicago Press, Chicago, 309-354.

Ben-Israel A., T.N.E. Greville (2003) Generalized inverses Theory and applications, 2nd ed., Springer Verlag, New York.

Bertsimas, D. and J.N. Tsitsiklis (1997) Introduction to Linear Optimization. Athena Scientific, Nashua.

Bikker, R.P. and S. Buijtenhek (2006) *Alignment of Quarterly Sector Accounts to Annual Data*. Statistics Netherlands, http://www.cbs.nl/NR/rdonlyres/D918B487-45C7-4C3C-ACD0-0E1C86E6CAFA/0/Benchmarking_QSA.pdf. (accessed May 2018).

Bikker, R.P., J.A. Daalmans and N. Mushkudiani, N. (2013) Benchmarking large accounting frameworks: a generalized multivariate model. *Economic Systems Research*, 25, 390-408.

Bishop, Y., S. Fienberg, P. Holland (1975) Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge.

Bloem, A.M., R.J. Dippelsman and N.O. Maehle (2001) *Quarterly National Accounts, Manual–Concepts, Data Sources and Compilation*. IMF, Washington, D.C, http://www.imf.org/external/pubs/ft/qna/2000/Textbook/index.htm. (accessed May 2018).

Boonstra, H.J., C.J. De Blois and G.J. Linders (2010) *Macro-integration with inequality constraints an application to the integration of transport and trade statistics.* Discussion Paper 10006, Statistics Netherlands, The Hague/Heerlen, http://www.cbs.nl/NR/rdonlyres/05D02F6E-29AB-4A04-BB45-417B82C43DA1/0/201006x10pub.pdf. (accessed May 2018).

Boot, J.C.G., W. Feibes and J.H.C. Lisman (1967) Further methods of derivation of quarterly figures from annual data. *Applied Statistics*, 16, 65–75.

Bozik, J.E. and M.C. Otto (1988) *Benchmarking: Evaluating Methods that Preserve Month-to-Month Changes*. Statistical Research Report Series 88/07, U.S. Census, Bureau, Washington DC, http://www.census.gov/srd/papers/pdf/rr88-07.pdf. (accessed May 2018)

Brent, G. (2016) *Quadratic Optimisation for Table Balancing in Official Statistics*, Research Paper, Australian Bureau of Statistics (ABS).

Bruni, R. (2005) Error correction for massive data sets. *Optimization Methods and Software*, 20, 291-310.

Bruni, R. and G. Bianchi (2012) A Formal Procedure for Finding Contradictions into a Set of Rules. *Applied Mathematical Sciences*, 6, 6253 - 6271.

Byron, R.P. (1978) The Estimation of Large Social Account Matrices. *Journal of the Royal Statistical Society*, 141, 359-367.

8

C

Calzaroni, M. and A. Puggioni (1998) Evaluation and Analysis of the Quality of the National Accounts Aggregates. http://ec.europa.eu/eurostat/documents/64157/4374310/38-EVALUATION-ANALYSIS-QUALITY-

NATIONAL-ACCOUNTS-AGGREGATES-1998.pdf (accessed May 2018).

Canning P. and Z. Wang (2005) A Flexible Mathematical Programming Model To Estimate Interregional Input-Output Accounts. *Journal of Regional Science*, 45, 539-563.

Causey, B. and M.L. Trager (1981) *Derivation of Solution to the Benchmarking Problem: Trend Revision*. Unpublished research notes, U.S. Census Bureau, Washington D.C. Available as an appendix in Bozik and Otto (1988).

Chen, B. (2006) *Balanced System of Industry Accounts for the U.S. and Structural Distribution of Statistical Discrepancy*. Working Paper WP2006-8, Bureau of Economic Analysis, Washington DC, http://www.bea.gov/papers/pdf/reconciliation_wp.pdf (accessed May 2018)

Chen, B. and S. H. Andrews (2008), Research spotlight: an empirical review of methods for temporal distribution and interpolation in the national accounts. *Survey of Current Business*, 88, 31-37.

Chen, D., R.G. Batson and Y. Dang (2010) Applied Integer Programming; Modelling and Solution; John Wiley & Sons, Hoboken.

Chinneck, J.W. (1997) Finding a useful subset of constraints for analysis in an infeasible linear program. *INFORMS Journal on Computing*, 9, 164–174.

Chmeiss, A., V. Krawczyk, and L. Sais (2008) Redundancy in CSPs. In: Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008), IOS Press, Amsterdam.

Cholette, P.A. (1984) Adjusting sub-annual series to yearly benchmarks. *Survey Methodology*, 10, 35-49.

Cholette, P.A. and E.B. Dagum (1994) Benchmarking time series with autocorrelated survey errors. *International Statistical Review*, 62, 365-377.

Chong, K.P. and S.H. Zak (2008) An Introduction to Optimization. Series in Discrete Mathematics and Optimization, John Wiley & Sons, Hoboken.

Chow, G.C. and A.L. Lin (1971) Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53, 372-375.

Cox, L.H. (2003) On properties of multi-dimensional statistical tables. *Journal of Statistical Planning and Inference,* 117, 251–273.

Cuevas A., E.M. Quilis and A. Espase (2015) Quarterly Regional GDP Flash Estimates by Means of Benchmarking and Chain Linking. *Journal of Official Statistics*, 31, 627-647.

D

Daalmans, J.A. and T. Di Fonzo (2014) *Denton PFD and GRP benchmarking are friends. An empirical evaluation on Dutch Supply and Use Tables*. Paper presented to the IIOA-conference, Lisboa 14-18 July. https://www.iioa.org/conferences/22nd/papers/files/1666.pdf (accessed May 2018).

Daalmans, J.A. (2015) *Estimating Detailed Frequency Tables from Registers and Sample Surveys*. Discussion Paper, Statistics Netherlands, The Hague. https://www.cbs.nl/nl-nl/achtergrond/2015/07/estimating-detailed-frequency-tables-from-registers-and-sample-surveys (accessed May 2018).

Dagum, E.B. and P.A. Cholette (2006) Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series, Springer Verlag, New York.

Dash Optimization (2017) *Xpress-Optimizer Reference manual*, Dash Optimization, Warwickshire.

Delden, A. and S. Scholtus (2017) *Correspondence between survey and admin data on quarterly turnover*. Discussion Paper 2017-03, Statistics Netherlands, The Hague. https://www.cbs.nl/en-gb/background/2017/07/correspondence-between-survey-and-admin-data-on-quarterly-turnover (accessed May 2018).

Denton, F.T. (1971) Adjustment of monthly or quarterly series to annual totals: An Approach based on quadratic minimization. *Journal of the American Statistical Association*, 66, 333, 99-102.

Deville, J.C., C.E. Särndal (1992) Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association,* 87, 376-382.

De Beer, J., J. Raymer, R. van der Erf and L. van Wissen (2010) Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe. *European Journal of Population*, 26, 459–481.

De Jonge, E. and M. van der Loo (2015) *Editrules: R package for parsing and manipulating of edit rules and error localization*. R Package Version 2.9-0.

De Jonge, E. and M. Van der Loo (2018) *Validatetools, Reference manual*, R-package, Version 0.4.3. https://cran.r-project.org/web/packages/validatetools/validatetools.pdf (accessed May 2018).

De Waal, T. de, J. Pannekoek and S. Scholtus (2011) Handbook of statistical data editing and imputation. John Wiley & Sons. New York.

De Waal, T. de (2015) *General Approaches for Consistent Estimation based on Administrative Data and Surveys*, Discussion Paper, Statistics Netherlands, The Hague https://www.cbs.nl/nl-nl/achtergrond/2015/37/general-approaches-for-consistent-estimation-based-on-administrative-data-and-surveys (accessed May 2018).

De Waal, T. de (2016) Obtaining numerically consistent estimates from a mix of administrative data and surveys. S*tatistical Journal of the IAOS*, 32, 231-243.

Dillig, I., T. Dillig, and A. Aiken (2010) Small formulas for large programs: On-line constraint simplification in scalable static analysis. In: Proceedings of the 17th international conference on Static analysis (SAS'10) Springer-Verlag, Berlin. http://theory.stanford.edu/~aiken/publications/papers/sas10.pdf (accessed May 2018).

Di Fonzo, T. (1990) The Estimation of *M* Disaggregate Time Series when Contemporaneous and Temporal Aggregates are Known. *The review of Economics and Statistics*, 72, 178-182.

Di Fonzo, T. and M. Marini (2003) *Benchmarking systems of seasonally adjusted time series according to Denton's movement preservation principle*. Working Paper 2003.09, University of Padova, http://www.oecd.org/dataoecd/59/19/21778574.pdf. (accessed May 2018)

Di Fonzo, T. and M. Marini (2005) *Benchmarking a system of time series:Denton's movement preservation principle vs a data based procedure.* Eurostat Methodologies and Working Papers, Luxemburg. http://ec.europa.eu/eurostat/documents/3888793/5836401/KS-DT-05-008-EN.PDF/e0b1ee11-a7cd-4e68-8b08-58eddf5d46d7. (accessed May 2018)

Di Fonzo, T. and M. Marini (2011) Simultaneous and two-step reconciliation of systems of time series: methodological and practical issues, *Journal of the Royal Statistical Society C*, 60, 143–164.

Di Fonzo, T. and M. Marini (2012a) Benchmarking time series according to a growth rates preservation principle, *Journal of Economic and Social Measurement*, 37, 225-252.

Di Fonzo, T. and M. Marini (2012b) *On the Extrapolation with the Denton Proportional Benchmarking Method*, International Monetary Fund, Washington D.C. https://www.imf.org/external/pubs/ft/wp/2012/wp12169.pdf (accessed May 2018).

Di Fonzo, T. and M. Marini (2013), A Newton's method for benchmarking time series In: Grigoletto M, F. Lisi and S. Petrone (eds.) Complex models and computational methods in statistics. Springer, Berlin, 109–121.

Di Fonzo, T., B. Chen, T. Howells and M. Marini (2014) *The Statistical Reconciliation of Time Series Accounts after a Benchmark Revision*, BEA Working Papers, Bureau of Economic Analysis, Washington D.C. https://www.bea.gov/papers/pdf/BCHEN_WP_2014.pdf (accessed May 2018).

Di Fonzo, T. and M. Marini (2015) Reconciliation of systems of time series according to a growth rates preservation principle. S*tatistical Methods & Applications*, 24, 651-669.

Durbin, J. and B. Quenneville (1997) Benchmarking by state space models. *International Statistical Review*, 65, 23-48.

E

European Commission. Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses (2008) *Official Journal of the European Union*, L218, 14-20.

Eurostat (2009), ESS Handbook for Quality Reports. Eurostat Methodologies and Working Papers, Luxemburg.

Eurostat (2013) *Handbook on quarterly national accounts – 2013 edition*. http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-13-004 (accessed May 2018).

F

Felfernig, A., Zehentner, C. and P. Blazek. 2011. "CoreDiag: Eliminating redundancy in constraint sets." Proceedings of *22nd International Workshop on Principles of Diagnosis*, October 4-7, 2011, Murnau, Germany. Available at: http://www.ist.tugraz.at/felfernig/images/stories/home/dx_corediag.pdf (accessed March 2017).

Fellegi, I.P. and D. Holt (1976) A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*, 71, 17-35.

Fernández, R.B. (1981) A Methodological Note on the Estimation of Time Series. *The review of Economics and Statistics*, 63, 471-476.

FICO (2009) Xpress-Optimizer Reference manual, http://www.fico.com/en/Products/DMTools/xpress-overview/Pages/Xpress-Documentation.aspx. (accessed May 2018)

Federal Committee on Statistical Methodology (2001) *Measuring and reporting sources of error in surveys*. Statistical Policy Working Papers, no. 31, Office of Management and Budget, Washington D.C, https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp31.pdf (accessed May 2018).

Fisher, I. (1922) The Making of Index Numbers, Houghton-Mifflin, Boston.

G

Gurobi (2016) *Gurobi Optimizer Reference Manual*, Gurobi Optimization Inc, Houston.

H

Harvill Hood, C.C. (2005). *An empirical comparison of methods for benchmarking seasonally adjusted series to annual totals*. Paper presented at the Workshop on Frontiers in Benchmarking Techniques and their Applications to Official Statistics, Luxembourg, 7-8 April 2005, https://www.census.gov/ts/papers/chood_asa2005.pdf (accessed May 2018).

Helfand, S.D., N.J. Monsour and M.L. Trager (1977) Historical Revision of Current Business Survey Estimates, In: Proceedings of the Business and Economic Statistics Section, American Statistical Association, Washington, 246–250.

Hillier, F.S. and G.J. Lieberman (2008) Introduction to Operations Research, McGraw-Hill, New York.

Hillmer, S.C. and A. Trabelsi (1987) Benchmarking of economic time series. *Journal of the American Statistical Association*, 82, 1064 – 1071.

Hooker, J. (2000). Logic-Based Methods for Optimization: Combining Optimization and Constraint Satisfaction. John Wiley & Sons, New York.

Houbiers, M., P. Knottnerus, A.H. Kroese, R.H. Renssen and V. Snijders (2003) Estimating Consistent Table Sets: Position Paper on Repeated Weighting. Discussion Paper, Statistics Netherlands, The Hague https://www.cbs.nl/nl-nl/achtergrond/2003/31/estimating-consistent-table-sets-position-paper-on-repeated-weighting (accessed May 2018).

Houbiers, M. (2004) Towards a Social Statistical Database and Unified Estimates at Statistics Netherlands. *Journal of Official Statistics,* 20, 55-75.

I

IBM (2015) *ILOG CPLEX 12.6 User's Manual for CPLEX*. IBM corporation, New York.

K

Knottnerus, P., C van Duin (2006) Variances in Repeated Weighting with an Application to the Dutch Labour Force Survey. *Journal of Official Statistics*, 22, 565–584.

Konis, K. (2016) *lpSolveAPI: R Interface for lpsolve.* R package version 5.5.2.0. https://cran.r-project.org/web/packages/lpSolveAPI/index.html (accessed May 2018).

Kooiman, P., K. Zeelenberg and W.J. Keller (2003) De nauwkeurigheid van geïntegreerde gegevens (The accuracy of integrated statistics). In: Algera, S.A., P. van der Laan and J. Nobel (eds), Gedacht en gemeten, Statistics Netherlands, Voorburg / Heerlen, 111-120. http://www.cbs.nl/nl-nl/menu/methoden/research/strategisch-onderzoek/sociale-dynamiek-en-arbeidsmarkt/publicaties/2004-gedacht-en-gemeten-pub.htm

L

Little, R.J.A., M. Wu (1991) Models for Contingency Tables with known margins when Target and Sampled Populations Differ. *Journal of the American Statistical Association*, 86, 87-95.

M

Magnus, J.R., J.W. van Tongeren and A.F. de Vos (2000) National Accounts Estimation Using Indicators Analysis. *Review of Income and Wealth*, 46, 329-350.

Magnus, J.R. and D. Danilov (2008) On the estimation of a large sparse Bayesian system: The Snaer program. *Computational Statistics & Data Analysis*, 52, 4203-4224.

Monsour, N.J. and M.L. Trager (1979) *Revision and benchmarking of business time series*. American Statistical Association, Proceedings of the Business and Economic Statistics Section, 333-337.

Mushkudiani N, J.A. Daalmans and J. Pannekoek (2015) Reconciliation of labour market statistics using macro-integration. *Statistical Journal of the IAOS*, 31, 257-262.

N

Nicolardi, V. (2000) Balancing Large Accounting Systems: An Application to the 1992 Italian I-O Tables. Paper presented at the 13th International Conference on Input-Output Techniques, University of Macerata, Italy. https://www.iioa.org/conferences/13th/files/Nicolardi_BalancingLargeSystems.pdf (accessed May 2018).

Nieuwenbroek, N.J., R.H. Renssen, L. Hofman (2000) Towards a generalized weighting system. In: Proceedings of the Second International Conference on Establishment Surveys, American Statistical Association, Alexandria, 667-676.

Nocedal, J. and S.J. Wright (2006) Numerical Optimization, 2nd edition. Springer, New York.

O

Öhlén, S. (2006) Benchmarking and Seasonal Adjustment – A study of Swedish GDP. Paper presented at the Conference on Seasonality,

8

Seasonal Adjustment and their Implications for short-term Analysis and Forecasting, Eurostat, Luxembourg. http://ec.europa.eu/eurostat/documents/4578629/4579718/OHLEN-AB.pdf (accessed May 2018).

Özyurt, D.B., R.W. Pike (2004) Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes, *Computers & Chemical Engineering*, 28, 381-402.

P

Pannekoek, J., S. Scholtus and M. van der Loo (2013) Automated and manual data editing: a view on process design and methodology. *Journal of Official Statistics*, 29, 511–537.

Paulraj, S. and P. Sumathi. (2010) A Comparative Study of Redundant Constraints Identification Methods in Linear Programming Problems. *Mathematical Problems in Engineering*, Article ID 723402.

Piette, C. (2008) Let the solver deal with redundancy. In: Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'08) IEEE Computer Society, Washington D.C. https://hal.archives-ouvertes.fr/hal-00865304/document (accessed May 2018).

R

Renssen, R.H., N.J. Nieuwenbroek (1997) Aligning Estimates for Common Variables in two or more Sample Surveys. *Journal of the American Statistical Association*, 90, 368-374.

Rossi, N. (1982) A Note on the Estimation of Disaggregate Time Series When The Aggregate is known. *The review of Economics and Statistics*, 64, 4, 695-696.

S

Särndal, C.E., B. Swensson, J. Wretman (1992) Model assisted survey sampling, Springer Verlag, New York.

Schneider M.H. and S.A. Zenios (1990) A comparative study of Algorithms for Matrix Balancing. *Operations Research*, 38, 439-455.

Scholtus, S. (2013) Automatic editing with hard and soft constraints. *Survey Methodology*, 39, 59-89.

Schulte Nordholt, E., J. van Zeijl, L. Hoeksma (2014) The Dutch Census 2011: analysis and methodology, Statistics Netherlands, The Hague. https://www.cbs.nl/NR/rdonlyres/5FDCE1B4-0654-45DA-8D7E-807A0213DE66/0/2014b57pub.pdf (accessed May 2018).

Stadler, K., R. Wood, T. Bulavskaya, C.J. Södersten, M. Simas, S. Schmidt, A. Usubiaga, *et al.* (2018) EXIOBASE3—Developing a time series of detailed environmentally extended multi-regional input-output tables. *Journal of Industrial Ecology*

Stone, J.R.N., D.G. Champerowne and J.E. Meade (1942) The Precision of National Income Estimates. *Reviews of Economic Studies*, 9, 111-135.

T

Tamhane A.C. and R.S.H. Mar (1995) Data reconciliation and gross Error Detection in Chemnical Process Networks, *Technometrics*, 27, 409-422.

Temurshoev, U. (2012) *Entropy-Based Benchmarking Methods*. University of Groningen, Groningen. http://www.ggdc.net/publications/memorandum/gd122.pdf (accessed May 2018).

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer and G.J. de Vries (2015) An Illustrated User Guide to the World Input–Output Database: the Case of Global

Automotive Production, *Review of International Economics*, 23, 575–605

Titova, N., D. Findley and B.C. Monsell (2010) Comparing the Causey-Trager method to the multiplicative Cholette-Dagum regression-based method of benchmarking sub-annual data to annual benchmarks. In: Joint Statistical Meeting 2010 Proceedings, Business and Economic Statistics Section, 3007-3021.

Trager, M.L. (1982) *Derivation of Solution to the Benchmarking Problem: Relative Revision*. Unpublished research notes, U.S. Census Bureau, Washington D.C.. Available as an appendix in Bozik and Otto (1988).

U

United Nations (2017). *2016 International Trade Statistics Yearbook. Volume I Trade by countries*. United Nations, New York.

W

Weale M. (1985) Testing linear hypothesis on national accounts data. *Economic Modelling*, 2, 8-16.

Weale, M. (1992) Estimation of data measured with error and subject to linear restrictions. *Journal of Applied Econometrics*, 7, 167-174.

Wroe, D., P. Kenny, U. Rizki and I. Weerakkody (1999) *Reliability and quality indicators for National Accounts Aggregates.* ONS, London. http://ec.europa.eu/eurostat/documents/64157/4374310/40-RELIABILITY-AND-QUALITY-INDIC-NATIONAL-ACCOUNTS-AGGREGATES-1999.pdf/ (accessed May 2018).

Y

Yingfu, X., A. Lennmalm, D. Lennartsson and A. de Groote (2017) Uncertainty and automatic balancing of national accounts with a Swedish application, *Statistical Journal of the IAOS*, preprint, 1-7.

8

9. Summary

The purpose of this thesis is to improve the applicability of formal macro integration methods. Macro integration adjusts values at an aggregate level to satisfy a set of predefined consistency rules. The problem often applies to large accounting frameworks with large numbers of variables and consistency rules and with a large degree of interdependence between variables. Most macro integration methods are based on well explored quadratic optimization methods, for which efficient solution methods exist that are suitable for large-scale applications.

Chapters 2-4 consider so-called Denton methods for benchmarking. Benchmarking is a macro integration problem with a time component. The aim is to achieve consistency between time series that are produced at different frequencies, e.g. quarterly and annually. This means for instance removing discrepancies between annual values and the sums of four quarterly values. Usually, the high frequency data are adjusted to align with the low frequency data, while preserving as much as possible the short-term movements of the high frequency data. Benchmarking monthly or quarterly series to annual data is a common practice in many National Statistical Institutes. Several benchmarking methods are available in the literature. The Denton methods are especially popularly applied, mainly because of their simplicity.

Chapter 2 extends an existing multivariate Denton method. The new method combines several methodological features, like: linear constraints, ratio constraints, soft constraints, weights, and inequalities in one model. Therefore, a wide range of modelling possibilities is supported. As demonstrated in Chapter 2, these functionalities are indispensable for National Accounts reconciliation. Statistics Netherlands currently uses the extended model in the production of National Accounts tables.

Chapter 3 compares the Denton method with Growth Rates Preservation (GRP). It is often claimed that the latter method is grounded on an ideal movement preservation principle. Some works in the literature argue therefore that GRP should be preferred over Denton. We show however that there are important drawbacks to GRP, relevant for practical applications, that are not often mentioned in the literature. The first one is that GRP does not satisfy the time reversibility property, an often-used criterion from index number theory. According to this property it should not matter for the results whether forward or backward growth rates are preserved. That is, benchmarking an original time series, $t = 1,\ldots,n$, or a 'reversed' time series, $t = n,\ldots,1$ gives the same results. A second drawback of GRP methods relates to the singularity in its objective function. Complications of this are: avoidance of close to zero outcomes, irregular peaks in results and unnecessary sign changes after benchmarking. Because of these two problems and because Denton is easier to apply, Denton can be considered a better alternative than GRP for most applications.

Chapter 4 deals with an application of the Denton method, based on a so-called sequential estimation process. The sequential estimation of relatively short series is necessary for many practical applications because of the practical concerns of re-adjusting results that have already been published. It was demonstrated that Denton methods are not always appropriate for a sequential estimation process. Abrupt changes of benchmarking corrections can occur at the boundaries of the estimation intervals, which do not conform with Denton's movement preservation principle. Chapter 4 proposes solutions for sequential benchmarking problems. Empirical applications demonstrate that these solutions actually improve on an existing Denton method.

Chapter 5 examines a new application of macro integration techniques to the Dutch Population Census. The Dutch census is produced from a variety of data sources that are already available at Statistics Netherlands. Since some data sources do not cover the entire target population, census compilation partly relies on estimation. For the Dutch Census several detailed contingency tables need to be produced with common marginal totals in different tables. Because different tables are estimated from different sources, there is a risk that common marginal totals in different tables are not estimated the same. European legislations do not tolerate such inconsistencies. To solve this problem, a so-called repeated weighting method has been previously developed and applied to the latest two Dutch censuses. The estimation is however troublesome due to estimation problems that are inherent to repeated weighting. Chapter 5 proposes to use Stone's macro integration method as an alternative for repeated weighing. As before, the approach is based on quadratic programming. Different from the previous chapters, there is no time element involved. The starting point was to simultaneously estimate all cells of all tables, but the large problem size caused problems. As a solution, "Divide-and-Conquer" techniques have been proposed that break down a large problem into smaller sub problems, that can be solved with the same methods as the original problem. An application to Census data demonstrated that these techniques work to prevent estimation problems and that at an aggregate level the results are very similar to those obtained with repeated weighting.

Chapter 6 applies existing techniques for constraint simplification that are studied in Artificial Intelligence and Operational Research to the data editing problem in official statistics. Data editing is the process of checking and correcting information as provided by individual respondents of a survey. A large number of constraints needs to be handled in many data editing applications. This chapter shows that data editing can benefit from automated constraint simplification techniques. Performance can be improved, which broadens the scope of applicability of automatic data editing. Flaws in edit rule formulation may be detected, which improves the quality of automatic edited data. The results of this chapter are also very relevant to macro integration, because of the many constraints that are usually observed in macro integration problems.